

Agglomerasjon i Norge - Urban vekst i perioden 1815-2001

Joakim Vold

29.08.2001

Forord

Denne oppgaven er skrevet som en del av en mastergrad i Samfunnsøkonomi, under veiledning av Jørn Rattsø.

Jeg vil gjerne takke Jørn Rattsø, som en sentral støttespiller, samt min familie og forlovede som ga meg både rom og støtte ut ifra hva jeg hadde behov for.

Oslo, 29. August 2011

Joakim Vold

Innhold

1	Innledning	1
2	Teori	2
2.1	Gibrats Lov	3
2.2	Zipfs Lov	10
3	Tidligere relevante analyser	17
3.1	Kanadiske byer	17
3.2	Urban Evolusjon av Amerikanske Byer	19
4	Data	23
4.1	Datasett	23
4.2	Deskriptiv statistikk	26
5	Metode	30
6	Analyse av Gibrats lov	34
7	Analyse av Zipfs Lov	39
8	Konklusjon	44
A	Tabeller	48
A.1	Tabeller for deskriptiv statistikk	48
A.2	Tabeller for økonometriske spesifikasjoner	50
B	Regresjons Resultater	51
B.1	Gibrat estimasjoner	51
B.2	Zipf estimasjoner	53

Figurer

1	Tetthetsdiagram. Y-aksen er varians, X-aksen er $\ln S_{it}$	9
2	Optimal Zipf. Y-aksen er $\ln(\text{Rang})$, X-aksen er $\ln(S_{it})$	10
3	Kernel regresjon med fast vekstsnitt	36
4	Kernel regresjon med vekstsnitt som varierer hvert tiår	37

Tabeller

1	1815	48
2	1875	48
3	1910	49
4	1950	49
5	2001	49
6	Between Effects estimasjon	50
7	Random Effects	51
8	Robust Random Effects	51
9	Fixed Effects	51
10	Robust Fixed Effects	52
11	BE estimasjon	53
12	t-statistikk for between effects estimasjon	54
13	Robust(Monte Carlo) BE estimasjon	54
14	t-statistikk for robust between effects estimasjon	55
15	Hill estimasjon	55
16	t-statistikk for Hill estimatorene	56
17	Robust Hill estimasjon	56
18	t-statistikk for de robust (Monte Carlo) Hill estimatorene	57
19	Robust Between Effects estimator	58
20	Robust Hill Estimator, total	58

1 Innledning

I 2001 hadde Norge 434 forskjellige kommuner fordelt over 19 forskjellige fylker. I denne fordelingen har Norge 14 urbane områder av forskjellige størrelser, som hovedsakelig består av en bykommune og nærliggende områder. Disse urbane områdene igjen er resultatet av agglomerasjonsmekanismer, som blir bestemt av positive og negative eksternaliteter. Det merkverdige er hvordan systemet av de urbane områdene er utformet. Hierarkiet som eksisterer blant de forskjellige byene er enkel å se, men vanskelig å forklare.

Denne oppgaven baserer seg på hvorvidt de norske urbane områdene følger kraftlovene som eksisterer i andre deler av verden. Fra Black og Henderson (Black and Henderson, 2003) har vi en analyse av de amerikanske urbane områdene, og hvorvidt distribusjonen av disse følger Zipfs lov. I tillegg kan man se på Clark og Stabler (Clarke and Stabler, 1991) som undersøker hvordan kanadiske byer følger Gibrats lov. Kanadiske byer er et godt utgangspunkt for analyse av Gibrats lov og norske byer, ettersom de urbane områdene er tilsvarende små i begge land. Dermed støter de kanadiske byene på de samme problemer som vil kunne oppstå ved norske data.

Vi vil i denne oppgaven undersøke spesifikt om hvorvidt norske data er underlagt Zipfs- og Gibrats lov, og om dette kan statistisk bevises. Teorien for Zipfs lov er hentet ut fra arbeidet til Xavier Gabaix (1999a; 1999b), men også arbeidet til for eksempel Tong Soo (2002) og Gabaix and M. Ioannides (2003). Gibrats lov henter vi teorien hovedsakelig ut ifra arbeidet til Eeckhout (2004), men også Lanaspá and Sanz (2010).

Vi vil også se på definisjonsspørsmålet rundt hva en by, eller et urbant område, kan være. Dette er nesten like viktig som selve teorien siden dette vil påvirke de endelige resultatene vi sitter igjen med. Mye av problematikken tar Black og Henderson (2003) for seg, og vi vil ta dette i betraktning når vi tar endelige avgjørelser.

Estimatene vil produseres ved en rekke paneldatametoder, men vil også se på ikke-parametrisk analyse for Gibrats lov. Den parametriske analysen vil testes ved bruk av standard t-tester, men vi vil også ta i bruk Hausman test for å se hvilken av de to metodene fixed og random effects er mest egnet. Statistikkprogrammet Stata 10 vil brukes for å gjennomføre estimasjonene og de statistiske testene oppgaven krever. Eventuelle tilleggsprogramvare som vi måtte ha behov for anskaffes også igjennom Stata.

2 Teori

Agglomerasjon er et enkelt konsept å forstå. Konsentrasjonen av mennesker innenfor et spesifikt område, gjerne en by eller et urbant område, forekom i sine tidligste former i områder som hadde geografiske fordeler. Det vil si sjøer, elver, kyststriper eller fjell, som ville gi fordeler innenfor matsanking (tilbake til jeger og samler formen for samfunn) eller handel og produksjon. Nå til dags kan en argumentere for at det er de økonomiske faktorene som hovedsakelig bidrar til avgjørelsen om hvor man skal bosette seg, og at geografisk beliggenhet er mindre viktig.

En annen måte å definere agglomerasjon på er ved fordelene som anskaffes når bedrifter eller individer samles i et mindre område. Fordelene i seg selv oppstår i all hovedsak fra kostnadsreduksjoner: reduserte fraktkostnader tilknyttet varer, individer eller ideer. Et eksempel på dette vil være Santa Clara Valley i California, bedre kjent som Silicon Valley, som hadde sin begynnelse ved byggingen av Stanford University. Universitetet ble grunnlagt blant åpne landskap, men ettersom begavede studenter, først og fremst Frederick Terman (L. Glaeser, 2011), begynte å bygge industriparker i nærheten endret dette seg til det som i dag er et sentralt område for teknologi og innovasjon. Men reduserte transportkostnader fører til et sentralt paradoks; ettersom viktigheten for agglomerasjon reduseres dersom transportkostnadene reduseres, vil man kunne anta at agglomerasjon mister sin relevans når kostnadene faller. Paradoksalt nok er ikke dette tilfellet, ettersom agglomerasjon fortsatt er et viktig konsept på tross av at handel foregår over store distanser. I tillegg til at verden blir komparativt mindre urbaniseres det sterkt innenfor enkeltnasjoner. Særlig synlig er dette i land som India og Kina, som opplever sterk økonomisk vekst, men også i byer som New York og London, som slet på 70-tallet (L. Glaeser, 2011). New York er et spesielt merkbart tilfelle, ettersom det på 70-tallet virket som om byen skulle bli slått konkurs (L. Glaeser, 2011).

I 2008 bodde ca. 3,5 milliarder mennesker i det som defineres som byer, mens antallet som bodde i bygdestrøk bestod av 3,4 milliarder mennesker. Med andre ord bor mer enn halvparten av verdens befolkning i byer, og FNs antakelser er at denne andelen vil øke i den urbane populasjonens favør (pop, 2008).

Videre er det to kraftlover som tar for seg agglomerasjonsfenomenet, sett fra et befolkningsmessig perspektiv. Man har Gibrats lov, som ble formulert av Robert Gibrat i 1931, og Zipfs lov, opprinnelig oppdaget av Felix Auerbach [1913] men brukt videre av George K. Zipf i 1949.

2.1 Gibrats Lov

Gibrats lov kan forenkles til ett grunnprinsipp: veksten på et urbant område er uavhengig av størrelsen. Med andre ord vil den variere mellom hvert urbane område, og det vil ikke eksistere noe systematisk mønster. Dette er fordi den stokastiske prosessen er lik for alle byer (Eeckhout, 2004). Så ideelt sett vil alle byene ha en vekst lik snittet. Uheldigvis kommer ikke dette uten komplikasjoner, spesifikt for beskrivelsen av populasjonsmobilitet. Det er to empiriske lover eller regulariteter som bidrar til dette.

1. De største urbane områdene må tilfredsstill Zipfs Lov.

-Observert av Alexandre Le Maitre så tidlig som i 1682, men ikke formalisert før 1913 og deretter 1949.

2. Vekstraten til en befolkning avhenger ikke av størrelsen til det urbane området.

-Det vil med andre ord si at vekstraten varierer mellom hvert urbane område, men det er ikke noe systematisk mønster som avhenger av størrelsen, siden den stokastiske prosessen er lik for alle byene.

Den første loven sier at størrelsene på de største byene er avhengig av å ha en lognormal distribusjon, og satt i en graf bør de ha et stigningstall på tilnærmet -1. Dette gjelder i all hovedsak den øvre halen av byenes størrelsesfordeling, men vi vil gå dypere inn på selve loven i neste teoridel (Eeckhout, 2004).

Den andre loven er ikke stort mer komplisert enn den første. Gitt at alle byene har samme vekstrate, med en ukjent varians, sier loven at den minste byen vil ha samme gjennomsnittlige vekstrate som den største byen. Dette igjen tilsier, som regulariteten fastslår, at størrelsen til en by ikke vil ha en signifikant påvirkning på byens vekstprosess. Dette vil utdypes nærmere senere i dette delkapittelet (Eeckhout, 2004).

I forhold til estimatet om hvorvidt Gibrats lov gjelder for et datasett, har vi to aktuelle måter å analysere dette på. Man kan se på kort sikt, det vil si mellom to tidsobservasjoner, og man kan se på det på lengre sikt. I vårt tilfelle vil kort sikt tilsi tiårs-perioden mellom folketellinger, mens lengre sikt enten vil si hundre årsperioder eller alternativt hele tidsserien i datasettet (Lanaspa and Sanz, 2010). Fra dette punktet av vil S_{it} defineres som størrelsen på et urbant område i , som måles i antall mennesker over periode t . Distribusjonen er gitt ved:

$$\phi(\hat{\mu}, \hat{\sigma}) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \frac{\ln(S - \hat{\mu})^2}{\hat{\sigma}^2}\right\}$$

hvor $\hat{\mu}$ er forventet utvalgssnitt på logaritmisk form, mens $\hat{\sigma}$ er forventet standardavvik (Eeckhout, 2004). Disse er uavhengige av S_{it} .

Videre definerer vi g som vekstraten til region i . Vekstraten til en region blir dermed definert som $S_{it} = S_{it-1}(1 + g)$. Videre tar vi logaritmen av dette, som vil redusere forvridningene, legger til det faktum at vekstraten kan avhengige av den initielle størrelsen, og får den følgende vekstligningen (Lanaspa and Sanz, 2010):

$$\ln(S_{it}) - \ln(S_{it-1}) = \nu + \beta \ln(S_{it-1}) + u_{it} \quad (1)$$

u_{it} er gitt som en tilfeldig variabel som representerer tilfeldige befolkningssjokk som vil endre veksten i en by. Samtidig er ν gitt ved $\ln(1 + g)$, og vi antar at både ν og u_{it} er individuelle og uavhengig distribuert for alle de urbane områdene. Forventet verdi av sjokkene er lik null, mens variansen er gitt ved $Var(u_{it}) = \sigma^2 \quad \forall i, t$ (Lanaspa and Sanz, 2010). Dersom β er lik null vil Gibrats lov holde, og veksten vil være uavhengig av den initielle befolkningen (Lanaspa and Sanz, 2010). Som et resultat av dette kan vi bevise at verdien til en bys størrelse er avhengig av antallet perioder, vekstraten samt den initielle størrelsen. Dette tilsier (Lanaspa and Sanz, 2010):

$$E(\ln S_{it}) = \nu t + \ln(S_{i0}) \quad (2)$$

Samtidig har vi at variansen er gitt på en tilsvarende måte:

$$Var(\ln S_{it}) = t\sigma^2 \quad (3)$$

Som vi ser vil både forventet verdi -og dermed også gjennomsnittet- og variansen variere over tid. Dermed kan vi se at proposjonal vekst fører til en lognormal distribusjon med standardavvik som øker i antall perioder. Dersom β er lik null vil urban vekst være parallell mellom områder (Lanaspa and Sanz, 2010), og vil ikke påvirkes av den initielle størrelsen. På den andre siden, hvis β skulle være statistisk signifikant forskjellig fra null vil den initielle størrelsen på det urbane området spille inn på vekstraten, og Gibrats lov vil falle (Lanaspa and Sanz, 2010). Dersom vi bruker Eaton-Ecksteins terminologi (Lanaspa and Sanz, 2010), vil tilfellet hvor β er lik null tilsi parallell vekst. Samsvarende har vi for $\beta \neq 0$, som tilsier divergent vekst mellom de urbane områdene dersom $\beta > 0$. Dersom vi har divergent vekst over lengre tid vil vi ende opp med en asymmetrisk fordeling hvor mindre urbane områder reduseres over tid komparativt med de største urbane områdene. På den andre siden av β , det vil si når β er mindre enn null, vil vi ha en konvergent vekst hvor de mindre områdene vil nærme seg de største over tid. Med andre ord vil en større initieell populasjon tilsi en lavere vekst (Lanaspa and Sanz, 2010).

Dersom $\beta \neq 0$, vil ligning 2 og 3 endres til henholdsvis (Lanaspa and Sanz, 2010):

$$E(\ln S_{it}) = \nu \frac{(\beta + 1)^t - 1}{\beta} + (\beta + 1)^t \ln(S_{i0}) \quad (4)$$

og

$$Var(\ln S_{it}) = \sigma^2 \frac{(\beta + 1)^{2t} - 1}{\beta^2 + 2\beta} \quad (5)$$

Dersom Gibrats lov holder, med β lik null, vil ligning 5 konvergere mot 3, og vi ser at det samme gjelder ligningene 4 og 2. For variansen utarter dette seg på følgende måte, gjennom bruken av L'Hôpitals regel (Lanaspa and Sanz, 2010):

$$\lim_{\beta \rightarrow 0} (\sigma^2 \frac{2t(\beta + 1)^{2t-1}}{2\beta + 2}) = \frac{2t\sigma^2}{2} = t\sigma^2 \quad (6)$$

På den andre siden, dersom $\beta > 0$ kan vi se endringen mellom variansen i de to tilfellene ved å subtrahere ligning 3 fra ligning 5. Dette gir oss (Lanaspa and Sanz, 2010):

$$3 - 5 = t\sigma^2 - \sigma^2 \frac{(\beta + 1)^{2t} - 1}{\beta^2 + 2\beta} = \frac{\sigma^2}{\beta(\beta + 2)} [t(\beta^2 + 2\beta) - (\beta + 1)^{2t} + 1] = \frac{\sigma^2}{\beta(\beta + 2)} [f(\beta)] \quad (7)$$

Denne ligningen er ikke definert med mindre $\beta > -1$, ettersom $f(\beta)$ vil være lik null dersom β er av verdien -1. I tillegg er funksjonen kun definert for $t > 0$. Siden vi i dette tilfellet ser på $\beta > 0$, vil ligning $\frac{\sigma^2}{\beta(\beta+2)} > 0$. Dersom $0 > \beta > -1$, så vil $\frac{\sigma^2}{\beta(\beta+2)} < 0$ (Lanaspa and Sanz, 2010).

Slik kan vi konkludere med at for å kunne si noe spesifikt om fortegnet i differansen mellom ligning 2 og ligning 4, må vi kunne si noe om hvordan $f(\beta)$ oppfører seg. Optimeringsproblemet for denne funksjonen, for å finne minimum- og maksimumspunktet, er gitt ved følgende førsteordensbetingelse (Lanaspa and Sanz, 2010):

$$\frac{df(\beta)}{d\beta} = f'(\beta) = 2t(\beta + 1 - (\beta + 1)^{2t-1}) = 0 \Rightarrow (\beta + 1)^{2t-1} = \beta + 1 \Rightarrow (\beta + 1)^{2t-2} = 1$$

Vi ser fra førsteordensbetingelsen at for at $f(\beta)$ skal være et minimum- eller maksimumspunkt, må $\beta = 0$ for å få $(\beta + 1)^{2t-2} = 1$. For å kunne vite om $\beta = 0$ er et maksimum eller minimum ser vi på andreordensbetingelsen (Lanaspa and Sanz, 2010):

$$\frac{d^2 f(\beta)}{d\beta^2} = f''(\beta) = 2t(1 - (2t - 1)(\beta + 1)^{2t-2}),$$

Dersom β er null, vil andreordensbetingelsen være gitt ved $f(\beta = 0) = 4t(1 - t) < 0$ så lenge $t > 1$. Dette betyr at funksjonen $f(\beta)$ er konkav og når sitt maksimale punkt i $\beta = 0$ og $t > 1$ (Lanaspa and Sanz, 2010).

Fra dette ser vi at fortegnet på differansen ved ligning 7 tilsier forskjellige vekstformer, gitt at antakelsene om $t > 1$ og $\beta > -1$ holder (Lanaspa and Sanz, 2010):

1. Ved $\beta > 0$ vil urban vekst være divergent. I tillegg vil variansen være større enn om Gibrats lov skulle holde. Med andre ord vil differansen være negativ.
2. Ved $\beta < 0$ vil den urbane veksten være konvergent, samtidig som at variansen vil være mindre enn i tilfellet der Gibrats lov holder. Med andre ord vil differansen være positiv.
3. Dersom $\beta = 0$ vil Gibrats lov holde og differansen i ligning 7 vil være lik null.

For å kunne se på lenger sikt, må vi endre størrelsesbetegnelsen. Der vi tidligere hadde befolkningsstørrelse vil vi nå gå over til relativ urban størrelse. Dette defineres som (Lanaspa and Sanz, 2010):

$$s_{it} = \frac{S_{it}}{\bar{S}_t} = \frac{S_{it}}{\frac{1}{N} \sum_{i=1}^N S_{it}}$$

Perspektivet er nødvendig for å kunne se på steady-state distribusjonen til de urbane områdene. Man vil her kombinere den parametriske- med den ikke-parametriske analysemetoden for å kunne analysere hvordan veksten og variansen i veksten endres i forhold til bystørrelsen.

Parametrisk analyse kan brukes til å oppdage hvorvidt Gibrats lov opptrer i dataene. Dette kan også sees på som tilsvarende med å teste etter enhetsrøtter (Clarke and Stabler, 1991). Problemet med denne framgangsmåten er måten dataene vil opptre på (Gabaix and M. Ioannides, 2003). Observasjonene repeteres kun hvert tiår eller sjeldnere, der det optimale ville vært årlige observasjoner. Vi har derfor kun nitten tidsobservasjoner. I tillegg vil det eksistere avhengighet i tverrsnittsdataene mellom byene, som igjen vil føre til estimater som ikke er særlig robuste (Lanaspa and Sanz, 2010).

Ved bruken av Peserans test etter enhetsrøtter, som er basert på t-ratioen til OSL-estimatet av b_i , kan vi bruke følgende tverrsnitts-utvidede regresjon (Lanaspa and Sanz, 2010):

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c \bar{y}_{t-1} + d_i \Delta \bar{y}_t + e_{it} \quad (8)$$

hvor y_{it} er lik $\ln s_{it}$, eller logaritmen av den relative størrelsen. Regresjonen er en standard Dickey-Fuller regresjon som utvides med snittet av de tidsforsinkede variablene, og første

2 TEORI

differansen til den avhengige variabelen. Dette gjøres for å asymptotisk filtrere ut påvirkningen av de uobserverbare fellestrekkene, samtidig som at vi ønsker å legge til rette for Pesarans test (Lanaspa and Sanz, 2010).

For å teste etter enhetsrøtter hos den naturlige logaritmen setter vi opp en nullhypotese: Alle seriene er ikke-stasjonære, og vi antar at Peresans tverrsnittsutvidede Dickey-Fuller er konsistent gitt alternativet om at kun en andel av seriene er stasjonære. Ett argument mot å bruke Peresans test er at den ikke er designet for få temporære observasjoner (Lanaspa and Sanz, 2010), men siden vi kun ser på de største ti byene veier det opp mot den reduserte mengden tidsobservasjoner.

Den enkleste modellen testen følger, som også kan brukes for å illustrere testen, kan gis som (Peseran, 2007):

$$y_{it} = (1 - \phi_i)\mu_i + \phi_i y_{i,t-1} + u_{it} \quad i = 1, \dots, N; t = 1, \dots, T \quad (9)$$

Som vi ser er denne modellen autoregressiv, hvor y_{it} er observasjonen for individ i ved tid t . Feiltermen vil være gitt ved en enkelt faktor og en tilfeldig feilfaktor (Peseran, 2007):

$$u_{it} = \gamma_i f_t + \epsilon_{it} \quad (10)$$

Her er f_t den uobserverte effekten som er felles for alle ledd, og ϵ_{it} er det idiosynkratiske feilleddet. For å tilpasse denne modellen til en 'standard' modell, skriver vi ligningene 9 og 10 som:

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i f_t + \epsilon_{it} \quad (11)$$

der α_i vil være gitt ved $(1 - \phi_i)\mu_i$, β_i er gitt ved $-(1 - \phi_i)$ og $\Delta y_{i,t}$ vil være gitt ved $y_{it} - y_{i,t-1}$. Enhetsrot hypotesen, som er gitt ved $\phi = 1$, kan beskrives som:

$$H_0 : \beta_i = 0 \forall i \quad (12)$$

Alternativhypotesen kan gis som $\beta_i < 0$ for $i = 1, 2, \dots, N_1$, og $\beta_i = 0$ for $i = N_1 + 1, N_1 + 2, \dots, N$. Brøken N_1/N angir her andelen stasjonære, individuelle prosesser. Vi kan anta at denne er en fast verdi mellom null og en når $N \rightarrow \infty$ (Peseran, 2007).

Det er tre antagelser rundt testen. Den første er at de idiosynkratiske sjokkene er uavhengig distribuert over både tid og individ, med snitt på null, varians lik σ_i^2 og et gitt fjerdeordens moment. Den andre antagelsen er at fellesfaktoren er ukorrelet i serier, har

snitt lik null, en konstant varians ved σ_f^2 og et gitt fjerde moments betingelse. Denne og forrige antagelse gir oss tilfellet hvor feilleddet 10 er ukorrelet i serier. Disse to antagelsene kan avslappes, og det vil ikke påvirke testen så lenge N og T går mot uendelig (Peseran, 2007). Den tredje og siste antagelsen er at ϵ_{it} , f_t og γ_i er uavhengig distribuert for alle i .

Den tverrsnittlige avhengigheten til feilleddet over i er vanlig i én-faktor modeller, til tross for at gyldigheten generelt sett kan kreve spesifikasjonen av mer enn én felles faktor 10 (Peseran, 2007).

Ikke-parametrisk analyse består av kernelregresjoner som er betinget på bystørrelse. Der ser man på spesifikasjonen:

$$g_i = m(s_i) + \epsilon_i \quad (13)$$

hvor g_i er den normaliserte vekstraten for en spesifikk region ($\ln s_{it} - \ln s_{it-1}$), og s_i er logaritmen til den relative størrelsen av urbant område i . $\hat{m}(s)$ blir i vårt tilfelle estimert som et lokalt snitt rundt punktet s , og utjevnes ved å bruke en kernel. For å kunne analysere effekten på lengre sikt bygger vi opp en samling av alle vekstratene mellom to påfølgende perioder. Metoden som deretter brukes er kjent som Nadaraya-Watson, og baserer seg på følgende uttrykk (Lanaspa and Sanz, 2010):

$$\hat{m}(s) = \frac{n^{-1} \sum_{i=1}^n K_h(s - s_i) g_i}{n^{-1} \sum_{i=1}^n K_h(s - s_i)} \quad (14)$$

K_h er avhengigheten av kernelen med båndbredde h , også kjent som en utjevningsparameter.

Vi kan bruke ligning 14 som utgangspunkt for å estimere variansen til vekstraten, ved å bruke Nadaraya-Watsen metoden:

$$\hat{\sigma}^2(s) = \frac{n^{-1} \sum_{i=1}^n K_h(s - s_i) (g_i - m(s))^2}{n^{-1} \sum_{i=1}^n K_h(s - s_i)} \quad (15)$$

Som vi ser fra ligningen over, samt ligning 14 er estimatoren særdeles sensitiv i forhold til ekstreme verdier. Et eksempel på dette vil være mindre urbane områder, som på grunn av størrelsen gjerne har mye større varians blant vekstratene. Dette er fordi en relativt liten økning i befolkningen representerer en større prosentvis vekst. (Lanaspa and Sanz, 2010)

2 TEORI

Dersom Gibrats lov stemmer vil veksten være uavhengig av størrelse i snitt og varians. Og ettersom vekstraten normaliseres, vil det ikke-parametriske estimatet være en tilnærmet rett linje rundt null-verdien. Samtidig vil estimatet av variansen være gitt ved en rett linje rundt verdien 1. Dette kan illustreres med følgende diagram: 2.1:



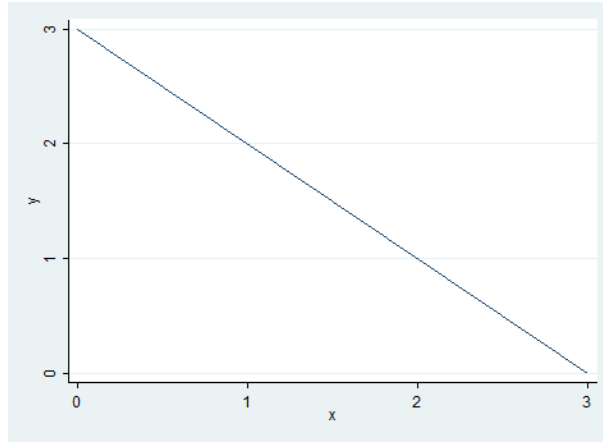
Figur 1: Tetthetsdiagram. Y-aksen er varians, X-aksen er $\ln S_{it}$

2.2 Zipfs Lov

Ifølge Xavier Gabaix (Gabaix, 1999b) er Zipfs lov ansett som en empirisk sett stabil lov, og blir ofte brukt som en streng betingelse for modellering av lokal vekst. Regularitetens grunnprinsipp går ut på å rangere urbane områder etter størrelse og deretter se etter et spesifikt mønster. Dette mønsteret er gitt ved en logaritmisk graf av rangen og befolkningen. Linjen som formes er en rett linje med en helning tilnærmet lik -1. Det er for så vidt ingen tautologi bak dette mønsteret (Gabaix, 1999b). Regresjonen som kjøres er gitt ved følgende ligning:

$$\ln Rank = \alpha - \zeta \ln Size \quad (16)$$

som er et uttrykk for Zipfs lov. Dersom \tilde{S} er størrelsen til et urbant område, gitt ved $Size$ i ligningen over, så vil Zipf's lov kunne skrives som $P(\tilde{S} > S) = \alpha S^{-\zeta}$, for en tilfeldig α , et vidt spenn av størrelser og eksponenten ζ . Eksponenten ζ er kjent som en paretoeksponent (Le Gallo and Chasco, 2008), som takket være konstruksjonsformen alltid vil være positiv (Gabaix, 1999b). Zipfs lov kan i tillegg uttrykkes som tilfellet hvor $\zeta = 1$, hvor den korresponderende tetthetsfunksjonen er gitt ved $p(S) = \zeta a S^{-\zeta-1}$, og den optimale modellen vil se ut som følger:



Figur 2: Optimal Zipf. Y-aksen er $\ln(Rang)$, X-aksen er $\ln(S_{it})$

I tillegg er det nyttig å se på den lokale eksponenten; $\zeta(S) = -Sp'(S)/p(S)$ hvor $p(S)$ er sannsynlighetsfordelingen til størrelsen S (Gabaix, 1999b).

Dersom vi ser på den grunnleggende modellen må vi komme med noen antakelser. Først og fremst antar vi at antall urbane områder er gitt, og at dette antallet er fastsatt. Videre antar vi at hvert byområde vokser med en arbitrær gjennomsnittlig vekstrate, med muligheten for vekstsjokk for hver enkelt by. Utover disse to antakelsene lar vi de urbane områdene utvikle seg fritt, for å kunne se på grensedistribusjonen. For å kunne se på en normalisert størrelse noterer vi med den normaliserte formen $S^i / \sum_{i=1}^N S^i$ (Gabaix, 1999b). Dette kan også beskrives som størrelsen til et urbant område, dividert på den totale urbane

befolkningen. Dermed har vi at S_t^i er den normaliserte størrelsen til et urbant område i periode t , som igjen vil tilsi at den gjennomsnittlige normaliserte bystørrelsen vil være konstant: $\sum_{i=1}^N S_t^i = 1$. Dette er fordi vekstprosessen er av formen $S_{t+1}^i = \gamma_{t+1}^i S_t^i$, hvor γ_{t+1}^i er en identisk og individuell distribuert tilfeldig variabel. Denne tilfeldige variabelen har en påfølgende distribusjon på $f(\gamma)$ (Gabaix, 1999b). Vi kan også definere $\gamma_{t+1}^i - 1$ som vekstraten til et urbant område, som igjen betyr at dersom den gjennomsnittlige normaliserte bystørrelsen skal forholde seg konstant må forventningsverdien til γ være gitt ved $E[\gamma] = 1$. Dette betyr med andre ord at gjennomsnittlig vekst for en by ikke kan eksistere. Eller, representert ved en funksjon, må være gitt som (Gabaix, 1999b):

$$\int_0^\infty \gamma f(\gamma) d\gamma = 1 \quad (17)$$

Den øvre halen av distribusjonen kan også betraktes, og vi betegner den øvre halen av urbane størrelser ved periode t som $G_t(S) = P(S_{t+1} > S)$. Bevegelsen til G_t kan presenteres i følgende ligninger (Gabaix, 1999b):

$$\begin{aligned} G_t(S) &= P(S_{t+1} > S) = P(\gamma_{t+1} S_t > S) = E[1_{S_t > S/\gamma_{t+1}}] \\ &= E[E[1_{S_t > S/\gamma_{t+1}} | \gamma_{t+1}]] = E[G_t(S/\gamma_{t+1})] \end{aligned}$$

som gir oss

$$\int_0^\infty G_t\left(\frac{S}{\gamma}\right) f(\gamma) d\gamma \quad (18)$$

La oss nå anta at vi har en steady-state prosess, slik at $G_t = G(S)$. Dette gir oss videre at:

$$G(S) = \int_0^\infty G\left(\frac{S}{\gamma}\right) f(\gamma) d\gamma \quad (19)$$

En distribusjons av typen $G(S) = aS^\zeta = a/S, \zeta = -1$ tilfredstiller steady-state distribusjonen i ligning 19, slik at man kan observere at Zipfs lov tilfredstiller steady-state distribusjonen (Gabaix, 1999b).

For å kunne bevise dette på en mer åpenbar måte ser vi på en modell som baserer seg på Gibrats lov. Vi antar at befolkningsvekst, som er en av de empirisk dominante faktorene (Gabaix, 1999b), i all hovedsak skjer gjennom migrasjon. Videre antar vi at det er overlappende generasjoner. I perioden en agent fødes, vil de avgjøre hvilket urbant område som gir dem størst nytte, gitt ved nyttefunksjonen $u(C) = a_{it}C$, og deretter migrere til dette området. Når denne avgjørelsen er tatt, vil de holde seg til det valgte urbane området. Dette er på grunn av antagelsen om at en eventuell kostnad ved å flytte vil være større enn nytteøkningen i neste periode. Dette er gitt en likevektssituasjon (Gabaix, 1999b).

Som vist fra nyttefunksjonen, er nytten avhengig av konsumet agenten kan utføre, samt fasilitetene til et område, og sjokk som skulle skje med fasilitetene kan sees på to måter. På den ene siden har vi sjokk som skjer på grunn av endringer i avgjørelsene et politisk regime tar. Dette kan komme i form av rentesatsendringer, kvalitetssikring ved samfunnsmessige institusjoner, eller avgjørelser som dekker kvaliteten til livsstilen hos beboerne. Sistnevnte kan både komme i form av endringer i tillatelser rundt forurensning, eller ved fritidsfasiliteter som kino, teater, eller lignende (Gabaix, 1999b).

Vi ser for oss to generasjoner som overlapper hverandre. Når en agent fødes, tar han det nevnte valget av tilholdsby. Mengden unge som migrerer til by i benevnes som N_i^Y , mens de av den eldre generasjonen som allerede oppholder seg i byen benevnes som N_i^O . Den eldre generasjonen har en sannsynlighet for å dø, gitt ved δ i hver periode. Avgjørelsen til agentene i denne modellen kan presenteres på følgende måte (Gabaix, 1999b):

$$\max a_{it}w_{it}, \quad (20)$$

hvor w_{it} er lønningen i by i ved periode t . I tillegg antar vi at det ikke er kapital inne i bildet for agentene. Ved en likevektssituasjon vil nyttetilpassede lønnsnivåer være like for alle byene. Ligningsformen blir dermed (Gabaix, 1999b):

$$a_{it}w_{it} = u_{it}, \quad (21)$$

som er en betingelse for likevekten. La oss nå anta at produksjonsteknologien har konstante skalaavkastninger, som gir oss følgende produksjonsfunksjon (Gabaix, 1999b):

$$F(N_i^O, N_i^Y) = N_i^O f\left(\frac{N_i^Y}{N_i^O}\right) \quad (22)$$

Vi har derfor at lønningene for den yngre generasjonen vil være gitt ved ligningen (Gabaix, 1999b)

$$w_i = w_i^Y = f'\left(\frac{N_i^Y}{N_i^O}\right) \quad (23)$$

Dersom vi kombinerer dette med likevektsbetingelsen ender vi opp med:

$$N_i^Y = N_i^O f'^{-1}(u_t/a_{it})$$

Gitt at befolkningsveksten til by i er gitt ved (Gabaix, 1999b):

$$\Delta N_{it} = N_{it}^Y - \delta N_{it}^O,$$

vil dette gi oss byens vekstligning:

$$\gamma_{it} = \frac{\Delta N_{it}}{N_{it}} = f'^{-1}\left(\frac{u_t}{a_{it}}\right) - \delta \quad (24)$$

På grunn av distribusjonen til fasilitets-variabelen, som er uavhengig av den initielle størrelsen, vil urbane vekstprosesser være identiske uansett størrelse.

Økonomisk intuisjon tilsier at byer som er relativt store vil ha en diversifisering av industriene som vil redusere effektene av eventuelle sjokk. Dette vil tilsi at variansen for en større by vil være lavere enn for mindre urbane områder (Gabaix, 1999b). Kontra til denne intuisjonen reduseres ikke variansen ved økende størrelse. Som vi ser er det en motsetning mellom empirien og intuisjonen (Gabaix, 1999b), som kan sees via vekstraten. Vekstraten kan sees på som summen av forskjellige enkeltsjokk, samt vekstsnittet($\bar{\gamma}$). Med andre ord:

$$\gamma^{it} = \bar{\gamma} + \gamma_{policy}^{it} + \gamma_{region}^{it} + \gamma_{industrier}^{it}$$

Når vi ser på variansen til denne ender vi opp med (Gabaix, 1999b):

$$\sigma^2(S) = \sigma_{policy}^2(S) + \sigma_{region}^2(S) + \sigma_{industrier}^2(S)$$

$\gamma_{policy(i)}^{it}$ vil være sjokkene som representerer endringer i fasilitetene, skattenivåene, eller andre former for politisk offentlige tjenester. Vi kan anta at den tilhørende variansen, σ_{policy}^2 , ikke avhenger av størrelsen. Det vil si at $\sigma_{policy}^2(S) = \sigma_{policy}^2$ (Gabaix, 1999b).

γ_{region}^{it} er effekten av regionsspesifikke sjokk. Dette kan være endringer i den økonomiske prestasjonen til region i , og eksempler på dette kan være en eller flere regioner som opplever sykluser som er kontra til det som oppleves i andre regioner. Vi kan også anta at variansen til regionen, i likhet med policy-variabelen, er uavhengig av størrelsen til byen. Dette kan vi anta siden byer som hører til samme region alle vil oppleve samme sjokk (Gabaix, 1999b).

$\gamma_{industrier}^{it}$ er effekten sjokk på industriene i et urbant område vil ha på populasjonsveksten. Som nevnt vil byens størrelsen spille inn på variansen til sjokkene det urbane området opplever. Dermed kan vi anta at $\sigma_{industrier}^2(S)$, ved hjelp av sentralgrenseteoremet, vil være gitt ved σ_{policy}^2/S . Dette endrer den totale variansen til følgende: (Gabaix, 1999b)

$$\sigma^2(S) = [\sigma_{policy}^2(S) + \sigma_{region}^2(S)] + \sigma_{industrier}^2(S)/S$$

Dersom vi ser på den øvre halen av distribusjonen vil utsagnet om at variansen ikke er

avhengig av størrelsen stemme, ettersom effekten av industriell diversifisering vil bli så redusert at den kan sees bort ifra. Dermed vil variansen være summen av to stedsinvariable varianser. Disse er variansen til sjokkene fra region, og variansen til sjokkene fra politisk atferd (Gabaix, 1999b).

Dette illustrerer essensen til mekanismen av Zipfs lov. På den andre siden er problemet med mekanismen at den ikke bringer en steady-state tilstand. For å oppnå dette trenger vi en annen mekanisme, som hindrer mindre urbane områder i å bli for små på grunn av for sterk befolkningsreduksjon (Gabaix, 1999b). Den enkleste måten å gjøre dette på er ved å bruke en random walk metode med en barriere. Dette vil forenkle analysen noe, i tillegg til at vi velger å se på modellen i kontinuerlig tid (Gabaix, 1999b).

ligningen som er utgangspunktet er:

$$\gamma_{it} = \frac{\Delta N_{it}}{N_{it}} = f'^{-1}\left(\frac{u_t}{a_{it}}\right) - \delta \quad (25)$$

Vi tar den kontinuerlige grenseverdien av ligning 25, som gir oss følgende ligning (Gabaix, 1999b):

$$\frac{dP_{it}}{P_{it}} = gdt + \sigma dB_{it} \quad (26)$$

B_{it} er en Brownsk bevegelse for en tilfeldig γ og σ , som avhenger av diskrete parametre. Disse parametrene er gitt ved u , a_{it} , og δ , som henholdsvis er nytten, fasilitetene, og sannsynligheten for død for den eldre generasjonen. I tillegg betrakter vi normaliserte størrelser, som er gitt ved $S_{it} = P_{it}/(\text{Urban populasjon forventet ved tid } t)$. Alternativt kan vi definere S_{it} som $S_{it} = P_{it}/(\text{Total forventet urban populasjon})$. Dette gjør at vi videre kan definere størrelsen som (Gabaix, 1999b):

$$S_{it} = \frac{P_{it}}{P_0 e^{gt}}$$

Samlet gir dette oss ligningen:

$$\frac{dS_{it}}{S_{it}} = \mu dt + \sigma dB_{it} \quad (27)$$

Her er forventet vekst gitt ved normaliserte størrelser. Med andre ord er μ gitt ved differansen mellom veksten til en by av en spesifikk størrelse og snittet til vekstraten (Gabaix, 1999b).

Grunnlaget for denne mekanismen ligger i det faktumet at uten mekanismen ville distri-

busjonen degenereres. Dette ville ført til at mindre byer ble for små, og dermed ville falle utenfor målingene.

I tilfellet med diskret tid gir bevegelsesligningen oss $S_{t+1} = \gamma_{t+1}S_t$, gitt at den forventede verdien av neste periodes vekstrate er lik 1. Bruker vi iterasjonsmetoden ender vi opp med $\ln(S_t) = \ln(S_0) + \sum_{\tau=1}^t \ln(\gamma_\tau)$. Videre vil loven om store tall vise oss at $\ln S_t/t \rightarrow E[\ln \gamma_{t+1}] < 0$, via Jensens ulikhet, som viser oss at $\ln(S_t)$ er veldig nære minus uendelig. Dermed kan vi teknisk sett si at sannsynligheten konvergerer mot null (Gabaix, 1999b).

I tilfellet hvor vi ser på kontinuerlig tid vil vi ende opp med $S_t = S_0 \exp\{-\sigma^2 t/2 + \sigma B_t\}$, som ikke har en steady-state distribusjon. Denne vil i all hovedsak være en lognormal distribusjon, hvor de fleste byer har en infinitesimal størrelse (Gabaix, 1999b).

En avart eller søsterlov av Zipfs lov er rang-størrelse-regelen. Denne ser på sammenhengen mellom størrelsen til et område, relativt med rangen den måtte ha. Konkret kan vi spesifisere dette som at følgende kraftlov gjelder: Størrelsen til en by ved rang i vil variere ettersom $1/i$ endres (Gabaix, 1999b), samtidig som at forholdet mellom størrelsen til urbane områder av forskjellig rang vil være gitt. Forholdet mellom den største og nest største byen vil være lik $1/2$, mens forholdet mellom rangeringene to og tre vil være gitt ved $2/3$. Generelt kan vi derfor si at forholdet mellom region av rang i og region av rang j er gitt ved R_j/R_i ($i < j$).

Med disse størrelsesratene kan vi sammenligne faktiske urbane mønstre med forventede Zipf-mønstre. Og gitt at Zipfs lov kan estimeres korrekt, og videre verifiseres, kan rang-størrelses-regelen på den andre siden bare approksimeres (Gabaix, 1999b). Derfor kan vi ikke forvente at forholdet mellom de forskjellige bystørrelsene faktisk vil være lik eller tilnærmet lik idealverdiene.

Dersom vi ordner et antall urbane områder etter størrelse, spesifikt ($S_{(1)} \geq S_{(2)} \cdots \geq S_{(N)}$), så vil snittet til S_j/S_i være gitt ved i/j . Dette er dersom vi antar at steady state -distribusjonen tilfredsstiller Zipfs lov, og vi har en eksponent på 1 (Gabaix, 1999b). Gjennomsnittet for det inverse forholdet (S_i/S_j) vil være lik $(j-1)/(i-1)$. Videre vil standardavvik være gitt ved $\sqrt{(1-i/j)(i/[j(j+1)])}$, og medianen vil være gitt ved den inverse generaliserte betafunksjonen, $B^{-1}(\frac{1}{2}, i, j-i)$ (Gabaix, 1999b).

Dette kan bevises ved å ta logaritmen av de ordnede bystørrelsene og se på effekten av normalisering. Ved normalisering vil de logaritmiske størrelsene følge en eksponensiell distribusjon under Zipfs lov:

$$P(s > t) = e^{-t} \quad , t \geq 0 \quad (28)$$

Som tidligere er s her den logaritmiske størrelsen til et urbant område ($\ln(S) = s$), og t

er realtiden. Videre kan vi, takket være Rényis representasjonsteorem, si at ved $i < j$ kan differansen $s_{(i)} - s_{(j)}$ skrives som følger (Reiss, 1989):

$$s_{(i)} - s_{(j)} = \sum_{k=1}^{j-1} \frac{x_k}{k} \quad (29)$$

Her er x_k uavhengige trekninger fra en eksponensiell distribusjon $P(x_k > x) = e^{-x}$ for $x \geq 0$. Dette tillater oss å beregne statistikken (Gabaix, 1999b):

$$E\left[\frac{S_{(j)}}{S_{(i)}}\right] = E[\exp(s_{(j)} - s_{(i)})] = E[\exp(-\sum_{k=i}^{j-1} \frac{x_k}{k})] = \prod_{k=i}^{j-1} E[\exp(-\frac{x_k}{k})] = \prod_{k=i}^{j-1} \frac{k}{k+2} = \frac{i}{j} \quad (30)$$

Samme framgangsmåte vil gi $E(S_{(i)}/S_{(j)}) = (j-1)/(i-1)$, som fører til beviset for snittet. Videre ser vi at selv om rang-størrelse-regelen stemmer, så gjelder dette kun ved forventede verdier og når forholdet er mellom en mindre by og en større by (*Mindre/Større*) (Gabaix, 1999b). Snittet til forholdet S_j/S_i vil være mindre enn 1 ved (i/j) , gitt at by i er større enn by j . Dersom dette forholde inverteres, (S_i/S_j) , vil dette ha en høyere verdi enn j/i takket være Jensens ulikhet -effekten (Gabaix, 1999b).

Dersom vi ser på verdien til standardavviket, så oppdager vi at disse vil være særdeles høye. Ved snittet til by 1 og 2, som er gitt ved 0.5, vil standardavviket være gitt ved $\sqrt{(1 - i/j)(i/[j(j+1)])}$, som gir oss verdien 0.2887. Dermed kan vi si at selv om to byer ligger ganske tett opptil hverandre vil ikke dette nødvendigvis motbevise Zipfs lov (Gabaix, 1999b).

3 Tidligere relevante analyser

Av tidligere relevante analyser, kan vi se på J. Stephen Clark og Jack C. Stabler sitt arbeid med å analysere Gibrats lov og veksten av kanadiske byer, samt Duncan Black og Vernon Hendersons arbeid med urban evolusjon av amerikanske byer og Zipfs lov.

3.1 Kanadiske byer

Clark og Stabler foretok i 1990 en studie av kanadiske byer og deres egenskaper. Hovedfokus var på 'diseconomies of scale', og hvorvidt dette kunne bevises gjennom Gibrats lov. Idéen var spesielt tilstedeværende på 1900-tallet. Akademikere argumenterte for at dersom byene var over en optimal størrelse ville de oppleve negative skalaavkastinger (Clarke and Stabler, 1991).

Testen Clark og Stabler gjennomfører er basert på Gibrats lov, og dersom denne holder vil den fortelle oss hvorvidt de urbane systemene har konstante skalaavkastninger, og dermed ikke er over den optimale størrelsen.

Clark og Stabler baserer seg på en annen modell for byvekst enn det vi har brukt, som er en variant av Ijiri-Simon/Vining-modellen (Clarke and Stabler, 1991). I denne modellen setter de opp en elementær vekstmodell, og dekomponerer så vekstraten til en tilfeldig faktor og en ikke-stokastisk komponent (Clarke and Stabler, 1991). Dette gir uttrykkene:

$$S_{it} = \rho_{it} S_{it-1} \quad (31)$$

$$\rho_{it} = \epsilon_{it} K_i \prod_{j=1}^n \rho_{it-j}^{\beta_{ij}} \quad (32)$$

som sammensatt og omformulert gir oss følgende ligning (Clarke and Stabler, 1991):

$$\Delta \ln(S_{it}) = c_i + \sum_{j=1}^n \beta_{ij} \Delta \ln(S_{it-j}) + u_{it} \quad (33)$$

Her er c_i lik $\ln(K_i)$, og u_{it} er lik $\ln(\epsilon_{it})$ (Clarke and Stabler, 1991). Ligning 33 er konsistent med Gibrats lov, ettersom veksten til by i er uavhengig av byens alder og størrelse (Clarke and Stabler, 1991). Dette testes ved å innføre et ledd som gir byens størrelse i forrige periode, og en tilsvarende parameter (Θ), og så estimere denne. Dersom parameteren Θ er lik null overholdes Gibrats lov og man har byer med konstant skalaavkastning. Skulle parameteren være mindre enn null vil de urbane områdene ha en vekstprosess som er avhengig av tidligere størrelser. Disse kan man omtale som stasjonære prosesser (Clarke

and Stabler, 1991). Testene man gjennomfører her vil basere seg på t-statistikk utvidet av Dickey og Fuller (Clarke and Stabler, 1991), som nevnt i teorikapittelet (Lanaspa and Sanz, 2010).

Datasettet til undersøkelsen utført av Clark og Stabler er det som gjør denne analysen relevant for vår oppgave. Det kanadiske urbane systemet minner i stor grad om det norske, ettersom det er få byer og disse er basert på definisjoner om urbane områder. Denne analysen ser på de største syv (Clarke and Stabler, 1991), der vi velger å fokusere på de største 14 byene. Videre har analysen 10 tidsobservasjoner, hvorpå vi har nesten doblet dette med 19 tidsobservasjoner. Grunnen til at vi omtaler disse som relativt like er på bakgrunn av problemene små datasett kan gi. Tester etter enhetsrøtter krever generelt sett litt større datasett, minimum 40 observasjoner per by, for å skaffe flere frihetsgrader.

Funnene til analysen er konsistent med andre undersøkelser. Det vil si at byene følger Gibrats lov i relativt stor grad. Et annet funn var at ordinær t-statistikk som antatt i større grad vil føre til at man forkaster hypotesene om at veksten til en by er uavhengig av tidligere størrelser. Dickey-Fuller utvidet t-statistikk gjør derimot ikke dette, noe som forteller oss at det er viktig å bruke korrekte kritiske verdier (Clarke and Stabler, 1991). Ut ifra disse funnene kunne ikke Stabler og Clark forkaste hypotesen om at de kanadiske byene erfarte konstante skalaavkastninger på tross av økende realinntekt med størrelse (Clarke and Stabler, 1991).

3.2 Urban Evolusjon av Amerikanske Byer

Black og Henderson utforsket i 2003 hvordan distribusjonen til det amerikanske urbane systemet var utformet. Blant det de så på, som er relevant for denne oppgaven, var hvorvidt størrelsesdistribusjonen fulgte Zipfs lov, og dermed også Gibrats lov.

Ideene som presenteres innledningsvis er standard for urban litteratur (Black and Henderson, 2003). Geografisk konsentrasjon av bedrifter i forhold til balansen mellom fordeler og ulempene ved agglomerasjon, som er med på å bestemme bystørrelser, samt sammensetningen av bedrifter i et urbant område (Black and Henderson, 2003). Den sistnevnte ideen kommer fra hvordan et urbant område påvirkes av sjokk, både den midlertidige og permanente formen. I tillegg vil informasjonen som spres via nærliggenhet -hovedsakelig takket være spesialisering av produksjon- være et annet aspekt av bedriftsammensetningen.

Ved midlertidige sjokk antar man at dette vil føre til økt mobilitet innad i størrelsesdistribusjonen. Permanente sjokk, endringer i nasjonalt produksjonskomposisjon, vil man på den andre siden anta at fører til endringer i selve distribusjonen.

En av undersøkelsene Black og Henderson foretar seg, er å se på hvordan urban evolusjon kan modelleres etter rang-størrelse-regelen og dermed paretodistribusjon. Videre etter dette ser de på alternative estimeringsmetoder, hovedsakelig ikke-parametrisk estimasjon, som faller utenfor perspektivet til denne oppgaven.

Datasettet Henderson og Black bruker er det amerikanske urbane systemet i perioden 1900 til 1990. De beskriver hvordan populasjonsendringen i denne perioden, på 225%, endret distribusjonen til fylkenes (amerikansk "County") befolkningsdistribusjon. Videre nevner de hvor stor andel av befolkningen som bor i urbaniserte områder, og hvor mye denne andelen har endret seg. Andelen startet på 40.5% i 1900, og avsluttet på 75.1% i 1990, med den sterkeste veksten fram til 1960 (Black and Henderson, 2003).

Selve datasettet er som nevnt fra perioden 1900 til 1990, og omhandler maksimalt 317 urbane områder, som består av 742 fylker. Men Black og Henderson har fire problemer med datasettet de spesifiserer. Det primære problemet er at definisjonen av fylkene har endret seg over tid. Akkurat som i vårt datasett har grenser endret seg, og fylker har både delt seg og slått seg sammen. Vi tok for oss problemet med å bruke Norsk samfunnsvitenskapelige datatjenestes beregninger. Black og Henderson velger å kombinere alle sammenslåinger og tilfeller hvor fylker deler seg, og skaffer seg en fellesnevner for fylkene. Dermed får de et konsistent sett med fylker over hele tidsperioden de ser på.

Det andre problemet deres var hvordan å definere urbane områder, slik at det blir korrekt over tid med endringene som nevnt. Analysen påpeker at definisjonen metropolområde, det vil være tilsvarende våre urbane områder, ikke oppstod før på 50-tallet (Black and Henderson, 2003). I tillegg har definisjonen gått gjennom betydelige endringer siden 50-

tallet, hovedsakelig fra politisk befolkningsmål til befolkningstetthet. Dette er et problem også fordi man ikke har befolkningstall tilbake i tid for inndelingen av de geografiske områdene slik de er satt per 1990, slik at man ikke kan gi konsistente definisjoner over tid. Black og Henderson løser dette ved å ta i bruk en metode fra Donal Bogue, hvor de følger utviklingen baklengs fra 1990 til 1900 og innfører moderne inndeling på historiske data.

Det tredje problemet ligger i hvordan man skal bestemme om et område kvalifiseres som et urbant område i hvert tiår. Man kan dermed se at problemet er sammensatt. På den ene siden ser man at dersom Black og Henderson hadde valgt et absolutt trunkeringspunkt, så ville dette ha ekskludert urbane områder som oppstod over tid (Black and Henderson, 2003). Argumentet mot det absolutte trunkeringspunktet er at teknologisk framgang har redusert relative kostnader og økt fordelene ved agglomerasjon. Dermed vil den gjennomsnittlige bystørrelsen øke over tid. Dette vil føre til at dersom man velger et trunkeringspunkt i forhold til de tidligste periodene, så vil dette videre føre til at flere steder blir kategorisert som urbane områder i senere tider (Black and Henderson, 2003). Dersom man setter dette punktet i henhold til senere tider, for eksempel 1980 eller 1990, vil dette igjen påvirke datautvalget ved å redusere antallet urbane områder tilgjengelig (Black and Henderson, 2003).

Black og Henderson velger da å innføre et trunkeringspunkt i hvert tiår, som representerer hvordan skillet mellom urbant område og tettsteder endres over tid. Hvert tiår gir et punkt lik den laveste befolkningen i et urbant område, dividert på gjennomsnittsbefolkningen i alle de urbane områdene.

Det avsluttende problemet til Black og Henderson er hvordan de kan definere størrelsen til de urbane områdene. På den ene siden kan man bruke urban populasjon, ut i fra tethetsdefinisjonen. Og på den andre siden kan man bruke total befolkning fra predefinerte urbane områder. Også her støter Black og Henderson på et definisjonsproblem, ettersom selve definisjonen på urban har endret seg over tid. De konkluderer med at siden de ikke kan rekonstruere definisjoner må de forholde seg til de kontemporære definisjonene.

Her har vi i denne oppgaven valgt å fokusere på den totale befolkningen innenfor definerte urbane områder, som gjennomført i Stortingsmelding nummer 31 fra 2002-2003 (stm, 2003) og forskningsrapporten om norske byregioner fra Agderforskning (Fosse et al., 2002).

Hoveddelen av analysen til Black og Henderson fokuserer på å karakterisere rommessig evolusjon i urbane systemer ved å analysere relativ størrelsesdistribusjon av urban befolkning, og hvordan disse endret seg over de 90 år Black og Henderson ser på (Black and Henderson, 2003). Den delen vi hovedsakelig hadde interesse av var der de brukte Zipfs lov på denne distribusjonen, spesifikt rang-størrelse-regelen. Utover dette ser de på Mar-

kovprosessene, mobilitet innad i systemet av urbane områders rangering, heterogenitet blant urbane steder, og urban spesialisering og transformasjon.

Black og Henderson begynner med å rangere byene som eksisterer i hvert tiår, fra den største med rang 1 til den minste med rang m . Dermed får man rang $R(n)$ for et urbant område av størrelse n . Distribusjonen antas å følge Zipfs lov, som gir en paretofordeling ved $R(n) = An^{-a}$, og kan empirisk estimeres med:

$$\ln(R(n_{it})) = \ln(A_t) - a_t \ln(n_{it}) + \epsilon_{it}, \quad i = 1, \dots, m_t \quad (34)$$

Rang-størrelse-regelen oppnås for a_t er lik 1. Black og Henderson bruker hypotesen om at paretoeksponenten har blitt redusert over tid i Amerika. Dette er for å vise til intuitjonen om at linjen dataene projiserer vil bli flattere over tid. Hypotesen insinuerer at den amerikanske befolkningen forflytter seg inn i urbane områder, som fører til økt urban konsentrasjon (Black and Henderson, 2003).

Resultatene av estimatene var generelt rundt 0.85, med noen variasjoner. Videre påpeker de at estimatet, som tidligere vist, er sensitivt i forhold til valget av datautvalg. Black og Henderson estimerer ligning 34 igjen, men denne gangen kun med den øverste tredjedelen av datasettet, for å emulere andre undersøkelser (Black and Henderson, 2003). Denne gangen startet estimatet tilnærmet lik 1, før det vokste de til komparativt høye nivåer i forhold til det opprinnelige datasettet.

Konklusjonen de tar ut ifra disse oppdagelsene er at forholdet til ligning 34 ikke er log-lineært definert. Black og Henderson ser så på en kvadratisk funksjon av typen:

$$\ln(R(n_{it})) = \ln(A_t) - a_t \ln(n_{it}) + b_t \ln(n_{it})^2 + \epsilon_{it}, \quad i = 1, \dots, n_t \quad (35)$$

Begrunnelsen for denne formuleringen er fra Gabaix (1999b), hvor de hintet om at dersom mindre og større byer er tilbøyelige for forskjellige sjokk så vil det eksistere et ikke-lineært forhold (Black and Henderson, 2003).

Fra Gabaix (1999b) har Black og Henderson også det faktum at Zipfs lov er et resultat av Gibrats lov, som vi har nevnt tidligere. Som beskrevet tidligere i oppgaven velger de også å teste etter enhetsrøtter, som fører til at de estimerer følgende ligning:

$$\ln(n_{it+1}) - \ln(S_{it}) = \alpha + \delta_t + \gamma \ln(N_{it}) + \epsilon_{it} \quad (36)$$

der δ_t er en vekstrate. Nullhypotesen Black og Henderson velger å teste, og som vi også velger å teste i resultatdelen, er gitt ved γ lik null. Hypotesen, som arbeidet påpeker, tilla-

ter ikke autoregressive tendenser hos modellen, og vi kan dermed bruke minste kvadraters metode (MKM) for å estimere modellen.

Resultatene som presenteres i henhold til dette gir oss at ligning 36 blir forkastet. Black og Henderson påpeker også et problem med datasettet, ettersom det hovedsakelig fokuserer på de sterkere byene, også kjent som vinnere (Black and Henderson, 2003). På grunn av dette vil det eksistere skjevheter i datasettet. De sterkeste byene er inkludert i datasettet, i motsetning til svakere byer, som ikke var i stand til å bli inkludert med varierende marginer.

Løsningen til Black og Henderson er å presisere at dersom en by kommer over trunkeringspunktet er den med i datasettet, selv om den skulle reduseres til under punktet. Men selv her vil Gibrats lov forkastes (Black and Henderson, 2003).

Vårt datasett har muligens det samme problemet. På den andre siden har vi predefinerte urbane områder, takket være Kommunal- og regionaldepartementet (stm, 2003). Datasettet vårt inkluderer dermed også det som Black og Henderson definerer som 'tapere', som igjen vil balansere informasjonen. Generelt vil da Zipf og Gibrats lov gi gode førsteordens beskrivelser av dataene, men dette fører til at Black og Henderson går videre til ikke-parametriske estimasjonsmetoder (Black and Henderson, 2003).

Konklusjonen til Black og Henderson viser til flere aspekter ved den urbane evolusjonen i Amerika. Blant dem er hvordan den urbane evolusjonen i seg selv involverer pågående økning i de urbane størrelsene, hovedsakelig takket være teknologiske fremskritt og hvordan systemet av byer gir en stabil størrelsesfordeling over tid. I tillegg gir dette en tendens der nye byer vokser meget raskt, mens byer generelt reduseres veldig langsomt. Videre er det en generell trend for økende urban konsentrasjon, med flere byer som havner blant de større kategoriene.

4 Data

4.1 Datasett

Befolkningstallene datasettet består av er plukket fra Statistisk sentralbyrås databaser, og baserer seg på Folke- og Boligtellinger i perioden 1815 - 2001. Her har man tatt i bruk Norsk samfunnsvitenskapelig datatjenestes tilpasning av befolkningsstrukturen i forhold til kommunegrensene som ble satt 1. januar 2002 for å skape et noe enklere datasett. Dette gjøres fordi norsk kommunestruktur har gått gjennom flere større endringer, hvor en de merkverdige kommunalinndelinger skjedde i perioden 1838 til 1998 (Juvkam, 1999).

Et eksempel på hvor komplisert endringene har vært igjennom årene er Tune. Opprinnelig var dette en kommune med 7020 innbyggere (Juvkam, 1999), men først en deling i 1839 -til Tune og Sarpsborg- og så enda en deling i 1861 som førte til kommunene Varteig og Tune, førte til en folkevekst på mindre enn tusen mennesker over 22 år. En tredje deling fant sted i 1911, noe som førte til at Tune og Rolvsøy oppstod. Tune, Vartieg, Sarpsborg og Skjeberg ble i 1992 slått sammen til kommunen Sarpsborg, samtidig som at deler av Varteig ble slått sammen med Rakkestad (Juvkam, 1999). Man bør spesifikt merke seg at de eldre folketellingene er noe usikre, ettersom tallene fra de tidligere periodene i all hovedsak er estimerer av enkelte kommuner som kan listes fra 2002 og utover men som ikke eksisterte på 1800-tallet (I. Severide and Longa, 2004).

Videre vil begrepet 'byregioner' brukes for områdene som skal analyseres. Byregionene består av en senterkommune, forstadskommuner, samt randsonerommuner (Fosse et al., 2002). Senterkommunen karakteriseres med en høy netto innpendling, lav befolkningsvekst, høye lønns- og utdanningsnivåer, bestemte levekårsproblemer og høye boligpriser. I større byer, for eksempel New York, Rio De Janeiro og lignende, kan levekårsproblemene ofte komme av mulighetene en by tilbyr (L. Glaeser, 2011).

Forstadskommuner består av kommuner hvor en høy andel av innbyggerene velger å pendle inn til senterkommunene (Fosse et al., 2002). Dette er den andre delen av bildet i forhold til senterkommunene, og vi kan si at forstadskommunene har en negativ netto innpendling. Videre vil forstadskommunene i likhet med senterkommunene ha høye lønns- og utdanningsnivåer, samt høye boligpriser, men lave levekårsproblemer og høy befolkningsvekst (Fosse et al., 2002).

Videre vil randsonerommunene være dårligere stilt enn forstadskommunene. Situasjonen er at deler av arbeidsstokken pendler inn til senterkommunen, lave boligpriser, lavere lønns- og utdanningsnivå, samt lav eller negativ befolkningsvekst. Det eksisterer også levekårsproblemer, men disse er gjerne av en annen art enn i senterkommunen (Fosse et al., 2002).

Den opprinnelige hensikten med en slik inndeling av de urbane områdene er å skaffe en felles plattform for analyse, slik at diskusjonen skal bli enklere mellom de forskjellige formene for analyse (Fosse et al., 2002). En annen årsak for denne inndelingen er av administrative hensyn. Kommuner er særegne administrative enheter, men samhandlingen med bykommuner vil endre fordelingen av bruken av offentlige tjenester (Fosse et al., 2002). Dette igjen gjenspeiler at samhandlingsmønsteret er en sentral del av bymønsteret, som vår type urban inndeling tar til rette for.

En annen årsak til denne formen for inndeling er at det strengt talt ikke eksisterer en kategorisering som tar for seg alle aspekter og utviklingmønstre av et byområde (Fosse et al., 2002). Samhandling mellom forskjellig kommuner, sosiale og romlige hensyn er med på å danne byområder, og de tidligere definisjonene som eksisterer tar ikke lenger hensyn til det komplekse bildet som eksisterer per i dag (Fosse et al., 2002).

Det som til dels er problematisk med datasettet er sammensettingen av kommuner som danner en byregion. Oslo-området er et særdeles vanskelig punkt, ettersom størrelsen på byregionen vil avhenge i særdeles stor grad av hvilke områder man tar med i beregningene. Som påpekt i rapporten om norske byregioner, (Fosse et al., 2002), vil folketallet i Osloregionen reduseres med 191,309 innbyggere dersom man ikke skulle inkludere Mosse- og Drammens-regionen. Og hvorvidt disse regionene inkluderes bedømmes av egenskapene de innehar. Blant disse er samspillet mellom Osloregionen og de mindre områdene, i form av beliggenheten, pendling og hvordan dette påvirker infrastrukturen, arbeidsmarkedene, og regionale funksjoner.

Det som taler for at disse to regionene, med omegnskommuner, skal tas med som en del av Oslo-regionen ligger i hvordan kollektivtransporten er utformet. Ettersom pendlingen inn til Oslo er relativ stor vil den kollektive trafikken tilpasses dette behovet. I tillegg vil dette endre på infrastrukturen, slik at områdenes tilkobling til Oslo tas i betraktning (Fosse et al., 2002).

På den andre siden av dette ser vi at både Mossregionen, og Drammensregionen har selvstendige arbeidsmarkeder (Fosse et al., 2002). Dette er i tillegg til at de begge er senterkommuner for omkringliggende kommuner, med positiv netto innpendling og som sees på som knutepunkter for nærliggende kommuner (Fosse et al., 2002). Drammen stiller spesielt sterkt her, ettersom det urbane området fungerer som et senter for fylket Buskerud (Fosse et al., 2002)

Avgjørelsen ble til slutt å inkludere disse mindre urbane områdene i Osloregionen på tross av argumentene for å separere dem. Dette gjøres hovedsakelig fordi sentrene ligger relativt tett samlet og samhandlingen mellom de tre områdene er for stor til å ignoreres.

Denne typen argumenter avgjør generelt hvilke kommuner som er en del av ulike urbane områder, men andre faktorer kan være med på å avgjøre. Bergen er et eksempel på dette,

hvor man i tillegg til de tidligere nevnte argumentene også kan ta i betraktning sendeflaten til regionavisen Bergens Tidene (Fosse et al., 2002).

Det største problemet med hvordan datasettet vårt er utformet dukker opp når man skal se på endringer over tid. Definisjonen av området omfattet av for eksempel Oslo omfatter er satt i 2001 av Kommunal- og Regionaldepartementet, men denne stemmer ikke nødvendigvis når en skal se på Oslo i begynnelsen av datasettet. Over tid har et urbant område vokst seg større, slik at det som muligens kunne sees på som en by i 1815 muligens kun kan sees på som en bydel i dag.

På den andre siden er de urbane strøkene definert ut ifra områder, og ikke trunkert etter befolkningstetthet og minimum befolkningstall. Metoden som Bogue bruker i sitt arbeid, (J. Bogue, 1953), er produsert ut ifra antakelsen om at et urbant område er definert av en størrelse og spesifikk befolkningstetthet. Dette gjør at denne metoden ikke fungerer i like stor grad for de innsamlede norske dataene. Dermed velger vi å bruke disse grensene, men har i bakhodet at det kan eksistere problemer med hvordan dataene er utformet.

4.2 Deskriptiv statistikk

Vi kan ganske enkelt se hvordan urban utvikling har endret seg opp igjennom årene. Fra kommunene som gradvis forsvinner i utkanten av Norge, til større byregioner som vokser jevnt over tid, vil relativt folketall og folkevekst være en god indikator på urbanisering. Framgangsmåten vår her vil være i henhold til innledningen fra Jan Eeckhouts 'Gibrat's Law For All Cities' (Eeckhout, 2004), hvor tabell 1 vil være gitt som mal.

Vi har valgt å fokusere på følgende årstall: 1815, 1870, 1910, 1950 og 2001. Det er flere forskjellige årsaker til dette.

1815 velges på grunn av at dette er den første av de jevnlige folketellingene, utført med ti års mellomrom, som foregår fram til 1875. Videre må det nevnes at 1815 også er den første folketellingen etter den norske grunnloven ble signert. Dermed ble ikke denne folketellingen styrt av den danske-norske administrasjonen, slik den var under de to forrige folketellingene i 1769 og 1801.

1875 er spesielt interessant ettersom det går en 15-års periode fra denne målingen fram til neste, som fant sted i 1890 ettersom det ble bestemt at folketellingene skulle utføres hvert hele tiår (Soltvedt, 2004). I tillegg er dette første folketellingen som koordinerer datoene for folketellingene, takket være et vedtak fra den internasjonale statistiske kongress (Soltvedt, 2004). Dette var i tillegg den første folketellingen som også inkluderte sjømenn.

1910 er den første folketellingen som ba de norske undersåttene om å oppgi hvilken nasjonalitet de hadde. Dette var en liten endring fra tidligere, hvor man kun oppga navn, alder og bosted. Videre må det nevnes at dette er første folketelling der Statistisk Sentralbyrå satt med et selvstendig faglig ansvar (Soltvedt, 2004).

1950 er den første regelmessige folketellingen etter andre verdenskrig, hvor de ville holde folketellinger hvert hele tiende år. I tillegg er dette første folketelling etter den overraskende sterke folkeveksten i 1946.

2001 var den siste folketellingen som baserte seg på et eget skjema (Soltvedt, 2004), og Statistisk Sentralbyrå bygger statistikk fra registre den dag i dag.

Tabellen i seg selv bygger på rang-størrelse-regelen, som er beskrevet i teorikapittelet. I hver tabell gir den første kolonnen rangen til regionen som angis i andre kolonne. Den tredje kolonnen viser folkemengden i forhold til den største regionen, hvorpå den fjerde kolonnen er folkemengden i forhold til snittet. Siste kolonne knyttes opp mot rang-størrelse-regelen, som er relativt viktig i forhold til Zipfs lov.

Regionene er som følger (Fosse et al., 2002):

- Oslo-regionen består av 46 kommuner, kategorisert til sentralkommune, forstads-kommunene og randsoneskommunene. I tillegg tas Mosse- og Drammensregionene med, ettersom disse sentrene har en høy konsentrasjon av innpendling til sentral-kommunen.
- Glomma-regionen, også betegnet som Nedre Glomma, består av 4 kommuner
- Tønsberg-regionene består av 7 kommuner
- Larvik-regionen er slått sammen med Sandefjordområdet og består av 3 kommuner.
- Grenlands-regionen består av 5 kommuner.
- Mjøsbyene består av 10 kommuner.
- Kristiansand-regionen består 10 kommuner.
- Stavanger-regionen består av 13 kommuner.
- Bergens-regionen består av 14 kommuner.
- Trondheims-regionen består av 10 kommuner.

Dersom vi ser på de individuelle tabellene for hver periode ser vi at Oslo-, Bergens- og Larvikregionen ikke endrer sin plassering i de utvalgte årstallene. Vi kan også se at Oslo over tid øker i størrelse relativt til de andre regionene. Den største endringen skjer i perioden 1815 til 1875, men den gjennomsnittlige veksten over de 60 årene er på 2,55%. Bergen er relativt stabil i forhold til Oslo, og reduseres ganske rolig mellom 1910 og 2001. I denne perioden ender Oslo opp med å gå fra 3.254 ganger så stor som Bergensregionen, til å ende opp som 3.738 ganger så stor. Dermed ser vi at Oslo i disse 81 årene opplever en liten relativ vekst i forhold til Bergen, hvis man sammenligner med 1800-talls perioden.

Ser man på rang-størrelse-regelen for 1815 1, så ser man at Oslo og Bergen er tilnærmet tilsvarende regelen med 0.465. Bergen til Mjøsbyene er 0.9, som er langt unna regelens $2/3$. Forholdet mellom Mjøsbyene og Trondheim er i samme båt som forrige tilfelle, med et forhold på 1. Igjen, dette er relativt langt unna rang-størrelse-regelen, som tilsier at dette forholdet skal være $3/4$.

Trondheim og Kristiansand-forholdet er også et stykke unna forholdsregelen på $4/5$.

Det er først på de lavere rangeringene at vi igjen nærmer oss rang-størrelse-regelen.

Forholdet mellom Grenland og Stavanger er på 0.935, noe som ligger litt over rang-størrelse-regelen, men man nærmer seg de $6/7$ -delene som regelen tilsier. Stavanger-Tønsberg-forholdet er på 0.962, hvor regelen sier $7/8$. Tilsvarende nære ligger forholdet

mellom Tønsberg og Glomma, på 0.93, som er relativt nære 8/9. Forholdet mellom Glomma og Larvik er igjen relativt nærme 9/10 forholdet som rang-størrelse-regelen tilsier.

I 1875 er det noen nevneverdige endringer 2. Blant annet faller Kristiansand fra rang 5 til rang 6, i tillegg til at Stavanger vokser seg fram til rang 5. Vi kan merke oss at Grenland faller fra rang 6 ned til rang 9, mens Glomma vokser opp til rang 7. Vi ser videre fra 1875 tabellen at mønsteret til dels gjentar seg selv fra 1815 tabellen. Verdt å merke seg er at forholdet mellom Oslo og Bergen er lavere enn i 1815. Forholdet er på 0,317, som er betydelig langt unna 1/2, som er forholdet rang-størrelse-regelen tilsier. Vi må ned til forholdet mellom Grenland og Larvik for å nærme oss rang-størrelse-regelen på 9/10.

Overgangen til 1910 varsler nok en gang om endringer 3. Som ved 1875-tellingen er forholdet mellom Bergen og Oslo noe svekket og langt fra rang-størrelse-regelen. Forholdet mellom Bergen og Trondheim ligger derimot ganske tett opptil rang-størrelse-regelen på 2/3, med en verdi på 0,656. Videre ser vi at regionene ligger relativt langt fra rang-størrelse-regelen. Trondheim har vokst seg til rang 3 -og dermed overgått Mjøsbyene, som falt til rang 4. Glommaregionen har vokst ytterligere, og ligger nå på rang 5. Kristiansand og Stavanger faller begge ned én rang, ned til henholdsvis 7. og 6. Tønsberg faller ned til rang 9, mens Larvik holder seg stabilt på rang 10.

Mellom 1910 og 1950 tellingene⁴ ser vi at de største endringene er at Stavanger øker sin størrelse, opp til rang 5. Samtidig faller Glomma ned én rang, Grenland går opp én rang, og Kristiansand faller til rang 8. Oslo, Bergen, Trondheim, Tønsberg og Larvik har samme rang som ved 1910 tellingen. I 1950 har vi igjen at Oslo og Bergen reduserer forholdet mellom seg, til 0,285. Nok en gang er dette langt ifra rang-størrelse-regelen. Videre ser vi at forholdet mellom Trondheim og Bergen ligger, som ved forrige tabell, tett opp til rang-størrelse-regelen. Det er kun 0,01 som skiller dette forholdet fra 2/6 som regelen tilsier. Ellers ligger forholdet mellom byene tett opp til hverandre, uten å nærme seg rang-størrelse-regelen. Det største avviket skjer mellom Grenland og Glomma, hvor forholdet er på 0,687 mens regelen tilsier et forhold på 6/7.

20015 byr på flere endringer fra 1950. Mest verdt å merke seg er at Stavanger klatrer fra rang 5 og ender opp med å bli Norges tredje største byregion. Trondheim og Mjøsbyene faller ned en rang hver, ned til henholdsvis fjerde og femte største urbane område. Kristiansand vokser seg til det sjette største urbane området, mens Glomma reduserer sin rang. Grenland reduseres relativt til de andre urbane områdene, slik at de ender opp som den niende største urbane regionen, mens Tønsberg vokser seg til å bli den åttende største urbane regionen. Rang-størrelse-regelen blir igjen utelatt til en viss grad. Det er først når vi nærmer oss Tønsberg at vi er tilnærmet lik rang-størrelse-regelen. Forholdet mellom Tønsberg og Glomma, som ifølge regelen skal være på 7/8 (0,875), er på 0,836. Utover dette varierer avviket fra 0,298 under regelen, til å ligge 0,116 over.

Som vi ser fra dette, blir Oslo regionen bare større og større i forhold til den gjennomsnittlige urbane størrelsen. Det er først i perioden mellom 1910 og 2001 at den relative veksten tilsynelatende roer seg ned. Den største veksten i Oslo som storby region, relativt til gjennomsnittet, skjer fram til 1910. Sammenligner vi dette med den urbane regionen Bergen, som har sterkest vekst mellom 1815 og 1875, eller Larvik, som snarere har en relativ reduksjon i folketallet relativt til snittet, så ser vi at den største urbane regionen stiller sterkest blant de ti største.

Dersom vi ser på den totale befolkningen i hver periode, så kan vi få et bilde av hvordan det urbane mønsteret har endret seg siden 1815. I 1815 så bodde det et minimum av 389 668 innenfor det som er definert som bygrenser(stm, 2003). Årsaken til at dette er et minimumstall, er på grunn av at man ikke har definert befolkningen i det urbane området Bodø. Men sammenlikner man dette med den totale befolkningen på 840 978 innbyggere, ser man at 46.3% av befolkningen bodde i sentrale strøk.

Ser man deretter på 2001, hvor den totale befolkningen var 4 520 829 innbyggere og 3 030 085 innbyggere bodde i urbane områder, ser man at hele 67% av befolkningen bor i byer. I 2002 er dette økt til 68% (Fosse et al., 2002).

Over en periode på 186 år, har andelen av befolkningen som bor i urbane strøk økt med 20.7 prosent poeng. Dermed ser vi at de urbane områdene har vokst særdeles sterkt over tid, både i andelen av befolkningen og i hvor mange mennesker som bosetter seg i urbane områder. Som nevnt i forrige delkapittel, vil definisjonen av de urbane områdene ha endret seg over tid. Dermed kan vi anta at andelen som bodde i det som tidligere var urbane strøk var mindre enn det antallet vi har ut ifra moderne definisjoner. Black og Henderson(Black and Henderson, 2003) hadde tilsvarende problemer. Men i motsetning til hos oss gikk bydefinisjonene deres på størrelser og befolkningstetthet. Spesifikt at et område måtte ha en befolkningstetthet på 150 innbyggere per kvadrat mile (2.58 kvadratkilometer), og et minimum av 50 000 innbyggere. Vår definisjon er som nevnt spesifikke områder knyttet til sentre, som er avhengig av et større netto pendling inn enn ut, og vil dermed kanskje være mer tilgivende ovenfor bruk av definisjonene for tidligere perioder.

5 Metode

Paneldataene organiseres etter en ID gitt til et urbant område, samt årstall. ID nummeret gis til hvert byområde, etter rekkefølgen de presenteres i hos Fosse, Karlsen, Magnusen og Cruickshank (Fosse et al., 2002) i kapittel 2. Årene er i henhold til Statistisk Sentralbyrås serie med folketellinger. Lovene vi modellerer etter, Gibrat og Zipfs ov, antar som vist tidligere at:

For Gibrats lov er antakelsene:

1. De største byene tilfredsstiller Zipfs lov
2. Den underliggende stokastiske prosessen i forhold til vekstrate er lik for alle byer.

For Zipfs lov er antakelsene:

1. Gibrats lov holder, slik at størrelsen ikke spiller inn på veksten.
2. Disribusjonen er i en steady state

De henholdsvis modellene som modelleres har utgangspunkt i ligning 1 for Gibrats lov, og ligning 16 for Zipfs lov. Gibrats lov vil gis ved ligningen den har som utgangspunkt, mens estimasjonen ved Zipfs lov vil utvides med et støyledd (ϵ_{it}). Dette gir oss:

$$\ln Rank = a - \zeta \ln Size + \epsilon_{it} \quad (37)$$

Paretoeksponenten kan sees på som en konvergensindikator, hvor en negativ endring vil føre til en divergent trend i bosetningsmønsteret. Inverst, så vil en økning i paretoeksponenten føre til en konvergent trend, og vi kan anta en agglomerasjon (Le Gallo and Chasco, 2008). Generelt sett kan vi si at det gjennomsnittlige estimatet av paretoeksponenten er underestimert, gitt at vi trekker n uavhengige og individuelle distribuerte (i.i.d) byer. For eksempel kan man trekke 100 i.i.d byer, og det gjennomsnittlige estimatet vil være gitt ved $\zeta = 0.94$, som betyr at en underestimering av eksponenten med 0.04. (Gabaix and M. Ioannides, 2003).

Videre, har vi skjevheter ved mindre utvalg av byer (Le Gallo and Chasco, 2008). Monte-Carlo undersøkelser har gitt en løsning, ved å ta en regresjon av Rang - $1/2$. Det vil si modellen er som følger:

$$\ln(Rang - \frac{1}{2}) = a - \zeta \ln(Size) + \epsilon_{it} \quad (38)$$

Fra Ibragimov og Gabaix (2006) (Gabaix and Ibragimov, 2006) har vi at standard estimasjon skjer ved at $(1/2)$, som er gitt fra parameteren ζ , er lik null. Det vil si at

$\ln(Rang - \xi) = \alpha - \zeta \ln(Size)$ hvor $\xi = 0$. Skjevhetene skjer ved at halen på distribusjonen er tykkere enn standard fordelingen, noe som gir en underestimering av ζ . Dette kan også representeres ved at standardavviket ved $\xi = 0$ er representert ved (Gabaix and Ibragimov, 2006):

$$\frac{2^{\frac{1}{2}}}{n} \zeta \quad (39)$$

I tillegg til dette kan OLS estimatene vi ønsker å gjennomføre lide under mangelen av en romlig autokorrelasjon hos residualene (Le Gallo and Chasco, 2008), som vil gi andre skjevheter i estimatene. Dette kan motvirkes ved å innføre nevnte residual, som korrigerer for dette, og å tillate forskjellige estimater av krysningsparameterene (Le Gallo and Chasco, 2008).

For å følge empirien, som gitt i Chasco og Gallo (2007) (Le Gallo and Chasco, 2008), så estimeres hver tids observasjon for seg. I denne oppgaven vil t være gitt ved 1815, 1825, 1835, 1845, 1865, 1875, 1890, 1900, 1910, 1920, 1930, 1946, 1950, 1960, 1970, 1980, 1990, og 2001. Her vil vi tillate en egen skjæringsparameter per år, slik at variasjon mellom perioder tillates.

For Gibrats lov vil vi i all hovedsak fokusere på bruken av fixed effects og random effects, med kernel-regresjon som en ikke-parametrisk analyse. De førstnevnte metodene vil utføres normalt, men også med robuste standardavvik. Modellen som estimeres vil være $\ln(S_{it}) - \ln(S_{it-1}) = \nu + \beta \ln(S_{it-1}) + u_{it}$.

Fixed effects blir her valgt, ettersom den ser på forskjellen mellom individene, og et tidssnitt. Dette skjer på grunn av bruken av informasjonen som en får av forskjellen mellom individene. Problemet med fixed effects metoden er at den ser på ν som et individsspesifikk ledd. Dette er ikke konsekvent med Gibrats lov, ettersom regulariteten krever lik vekst for alle urbane områder (Eeckhout, 2004). Til tross for dette ser estimasjonsmetoden bort fra det individsspesifikke leddet når den estimerer sammenhengen mellom vekst, i form av $\ln(S_{it}) - \ln(S_{it-1})$, og tidligere størrelser. Krysnings termen (α_i) vil estimeres uskjeve som $\hat{\alpha} = \sigma_{t=1}^T (1/T) (\ln(S_{it}) - \ln(S_{it-1})) - \beta \ln(\bar{S}_i)$. Dette medfører at statistikkens vi bruker for å teste null-hypotesen ikke nødvendigvis vil påvirkes i nevneverdig grad. På den andre siden må en ha i bakhodet at estimatene hovedsakelig bruker informasjonen som en får mellom individene, slik at det kan oppstå eventuelle skjevheter.

Random effects, som er den andre estimeringsformen vi bruker, vil se på feilleddet som et sammensatt element bestående av et tilfeldig individspesifikk ledd, og et ledd som beskriver plutselige sjokk. Det individsspesifikke leddet blir sett på som en tilfeldig faktor. Dette gjør at eventuelle faktorer som kan påvirke veksten, men som ikke tas med i resten av likningen, vil kunne oppsummeres uten å inkludere spesifikk høyreside-variable.

Formelt sett vil de to metodene estimere to forskjellig ting. Fixed effects vil estimere $E\{y_{it}|x_{it}, \alpha_i\} = x_{it}\beta + \alpha_i$, mens random effects hovedsakelig vil estimere $E\{y_{it}|x_{it}\} = \beta x_{it}$. Disse to vil generelt være identiske, dersom $E\{\alpha_i|x_{it}\} = 0$ Verbeek (2008).

Dermed vil vi ta i bruk begge estimasjonsmetodene, og så ta en Hausmans-spesifikasjonstest for å avgjøre om begge metodene er effektive, eller om det er korrelasjon mellom x_{it} og α_i . En Hausmans spesifikasjonstest er gitt som:

$$\xi_H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[\hat{V}\{\hat{\beta}_{FE}\} - \hat{V}\{\hat{\beta}_{RE}\}]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \quad (40)$$

Her vil nullhypotesen være gitt ved at det ikke er noen betydelig forskjell mellom de to estimatene, hvor fixed effects estimatet vil holde under null hypotesen og alternativhypotesen, mens random effects vil kun være konsistent under nullhypotesen. Statistikken som vi tester mot her er χ^2 fordelt med K frihetsgrader. Dersom vi ikke kan forkaste nullhypotesen, vil ikke variabelen være avhengig av det individsspesifikke leddet, hvilket vil føre til at begge estimatene er korrekte Verbeek (2008).

Feilleddet vil ha autokorrelative egenskaper, med mindre variansen til det individsspesifikke leddet er antatt å være 0. Dersom antagelsen om at den tidligere størrelsen ikke er korrelert med det sammensatte feilleddet, vil dette føre til at estimatene er konsistente og forventningsrette. Er dette ikke tilfellet, så vil det være forventningsskjevheter i standardfeilen, slik at vi må korrigere for den ukjente effekten Verbeek (2008). Den enkleste metoden vi kan bruke er å anvende standard minste-kvadraters-metode estimerer, fixed og random effects, og så justere standardfeilene. For oss vil dette si å estimere fixed og random effects med robuste standardfeil, som vi ser i tabellene 8 og 10.

Kernel regresjonen som vi vil bruke under undersøkelsen av gibrats lov vil være en Gaussisk kernel, siden gibrats lov hovedsakelig har en normalfordeling. Båndbredden beregnes automatisk fra Stata, og er gitt i figurene 6 og 6.

Når vi kommer til zipfs lov, vil ikke random effects eller fixed effects gi noen resultater i Stata. Fixed effects gir feilbeskjeden om at det ikke eksisterer nok observasjoner til å gjennomføre regresjonen, mens random effects ikke gir noen beskjeder. Det må merkes at dette gjelder når vi regresser per tidsobservasjon, slik at vi får tabellene 11, 13, 15, og 17. Vi testet derfor between effects metoden, som ga resultater, samt maximum likelihood metoden som også er kjent som Hill metoden Gabaix and M. Ioannides (2003).

Dersom det ikke eksisterer noen skjevheter, så vil between effects være identisk med minste-kvadraters metode. Dette fører til at disse estimatene er forventningsrette, og ikke innehar noen skjevheter. Dermed er dette et alternativ for fixed effects og random effects metodene Verbeek (2008).

Hill estimatoren, som Gabaix og Ioannides referer til, er en alternativ måte å estimere zipfs lov på. Hill er i essensen maximum likelihood estimasjon og vil vise til, dersom regulariteten er perfekt, særdeles gode estimater Gabaix and M. Ioannides (2003).

Dette gjør at vi vil bruke disse til å estimere zipfs lov med i analyse-delen av denne oppgaven.

Estimatet av Zipfs lov vil være av formen:

$$\ln R_{it} = \alpha - \zeta * \ln(S_{it})$$

Hvor R_{it} er rangen et urbant område(i) får i periode t, $\alpha = \ln(a)$ og er en parameter, S_{it} er folketallet i det urbane området i i periode t , og ζ er paretoparameteren. Vi tar da den nevnte between-effects-regresjonen på datasettet, rapportert ved hvert folketellingsår, og estimerer $\ln(R_{it}) = \hat{\alpha} + \hat{\zeta} * \ln(S_{it})$. Dette gir tabell 11. Goodness-of-fit, eller i hvor stor grad modellen forklarer virkeligheten, er gitt som R^2 i tabellen. Denne er lik $R^2_{between}$, ettersom dette er en between effects estimasjon og Stata ikke rapporterer noen verdi for R^2_{within} .

T-statistikken for estimatene av ζ i tabell 11 er gitt ved tabell 12.

Tar man et maximum likelihood estimat av de to alternativene, så får vi tabellene 15 og 17.

6 Analyse av Gibrats lov

Den parametriske modellen vil være av formen $\ln(S_{it}) - \ln(S_{it-1}) = \nu + \beta \ln(S_{it-1}) + u_{it}$, som insinuerer at størrelses endringen til et urbant område avhenger av vekstraten og av størrelsen i forrige periode. ν er, som nevnt tidligere, gitt ved $\ln(1 + g)$. Dersom vi tar en regresjon på dette med hensyn til ligningen over, vil vi anse ν som det konstante leddet som produseres ut fra regresjonsprogramvare.

Vi starter først med en ordinær lineær regresjon, ved en random effects estimasjon. Resultatet fra denne gis ved tabell 7.

Som vi ser er estimatet gitt som tilnærmet lik null, med en standard feil også tilnærmet lik null.

Nullhypotesen for testene etter Gibrats lov vil være gitt ved:

$$H_0 : \beta = 0 \quad (41)$$

Det vil si at vi ønsker å se hvorvidt modellen kun er avhengig av vekstraten.

Alternativhypotesen er gitt som:

$$H_1 : \beta \neq 0 \quad (42)$$

Med andre ord ser vi fra alternativhypotesen om veksten er avhengig av tidligere urbane størrelser. På denne måten kan vi i en vesentlig grad anslå hvorvidt Gibrats lov stemmer for Norge, dersom null hypotesen holder, eller om veksten faktisk er avhengig av størrelsen.

En standard t-test gir oss her en t-statistikk på 1.90, som er mindre enn den kritiske verdien på 1,96 ved 265 frihetsgrader. Dette betyr at vi ikke kan forkaste nullhypotesen til fordel for alternativhypotesen, og at β ikke er statistisk forskjellig fra null. Dette forteller oss at Gibrats lov stemmer for Norges urbane systemer.

Dersom vi ser på et estimat med robust standardavvik, gir dette et litt annet bilde. Her er resultatet fra regresjonen gitt ved tabell 8. Her ser vi at estimatene er like, men standard feilen er noe redusert. Dersom vi tar en t-test med disse estimatene, vil dette gi oss en t-statistikk på 2.48, som er nok til å forkaste nullhypotesen i 95% konfidensintervall med kritisk verdi på 1.96. Alternativt kan man si at β er statistisk forskjellig fra null. Dersom vi ser på et 99% konfidensintervall, er det ikke nok til å forkaste nullhypotesen med kritisk verdi på 2.576. Dette forteller oss at Gibrats lov ikke nødvendigvis gjelder for norske urbane mønstre, gitt at vi ser på 95% intervallet. Ved 99% vil forskjellen mellom

estimatet av avhengigheten mellom veksten og tidligere størrelser, og observert verdi, være statistisk insignifikant på 1% nivået.

Videre kan vi se på Fixed Effects, for å se om resultatet avhenger av estimasjonsmetoden. Her vil de ordinære estimasjonsresultatene kunne sees ved tabell 9, og de robuste estimatene vil kunne observeres ved tabell 8.

Igjen kan vi se at de estimerte verdiene er tilnærmet lik null, men at den robuste standard feilen er noe økt i forhold til den ordinære. Nullhypotesen er gitt ved $H_o : \beta = 0$, med alternativhypotesen gitt ved $H_o : \beta \neq 0$.

De statistiske testene viser her mer eller mindre det samme som ved random effects estimatene. For de ordinære estimatene vil t-testen gi oss en t-statistikk på 1.73, som er mindre enn den kritiske verdien på 1.96. Dermed så vil man kunne si at β ikke er statistisk forskjellig fra null, og at Norske urbane mønstre følger Gibrats lov.

De robuste estimatene gir oss en litt høyere t-statistikk. Med verdien 2.09 ligger dette over den kritiske verdien for 95% konfidensintervallet, men under den kritiske verdien for 99% konfidensintervallet. Dermed forkaster vi nullhypotesen for det førstnevnte intervallet, men beholder den under sistnevnte.

Fra den parametriske analysen kan vi si at dersom vi velger et stort nok konfidensintervall, her 99%, så vil det norske urbane systemet følge Gibrats lov. Med andre ord vil veksten være lik for alle byene, og det er kun tilfeldige sjokk som i lengden vil skille byene fra hverandre. Dersom vi skulle ha sett på R^2 for begge estimasjonsmetodene, så ser vi at denne er lik for begge modellene. Generelt ligger denne på 0.0125 for R^2_{within} , og 0.1651 for $R^2_{between}$. Dette henter om at modellen ikke har en stor goodness-of-fit i dette tilfellet, noe som henter om en relativ svak forklaringsevne når en tar med tidligere størrelser på byene.

Ved å ta en Hausman test mellom de to metodene, fixed og random effects, så vil vi få en χ^2 statistikk på 0.15. Dette betyr at det ikke er noen nevneverdig forskjell mellom de to estimatene til metodene, slik at vi ikke kan forkaste nullhypotesen. Nullhypotesen er, som nevnt i metode-delen, at det ikke er noen betydelig forskjell mellom metodene og deres estimater. Den kritiske verdien som eventuelt måtte passeres med en χ^2 med en frihetsgrad ville være 3.84 med et 95% konfidensintervall. Dermed kan de to metodene teoretisk sett brukes om hverandre.

En annen måte å se om hvorvidt Gibrats lov stemmer ovenfor er ved å utføre en ikke-parametrisk analyse. De enkleste metodene vi kan ta i bruk, er å se på fordelingen til vekstratene. Mer spesifikt vil vi se på hvordan fordelingen av vekstratene er i forhold til snittet. Og snittene kan igjen deles inn og produseres på to forskjellige måter.

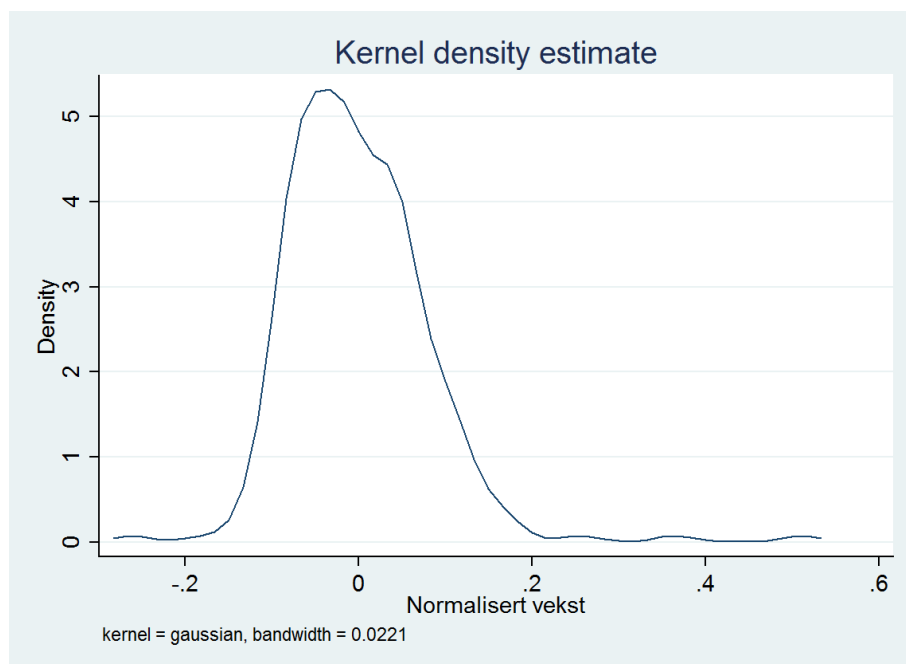
På den ene siden kan vi se på hvordan vekstratene divergerer fra et snitt som er likt over tid. Med andre ord kan vi se på et snitt som både spenner over tid, og individuelle observasjoner. På den andre siden kan vi se på hvordan veksten divergerer fra et snitt som er ti-års-spesifikt, som vil gi et bedre bilde av hvordan veksten utvikler seg over tid, men også hvordan veksten er i forhold til Gibrats lov.

Ved å normalisere vekstraten via differansen til vekstraten og gjennomsnittet for en by (G. Overman and Ioannides, 2000), kan vi se i hvor stor grad dataene eventuelt avviker fra Gibrats lov.

De to måtene å normalisere på har hvert sitt fortrinn. Ved å ta i bruk et totalsnitt, $(1/NT)\sum_{n=1,t=1815}^{T,N} g_{it}$, når man normaliserer, ser vi hvordan veksten avviker fra en normal som går over flere år. Vekstraten i seg selv blir beregnet fra:

$$g = \frac{S_{it} - S_{it-1}}{S_{it-1}} \quad (43)$$

Kernel regresjonen av den normaliserte vekstraten gir oss grafen 6, som viser oss hvordan fordelingen til den normaliserte vekstraten er.



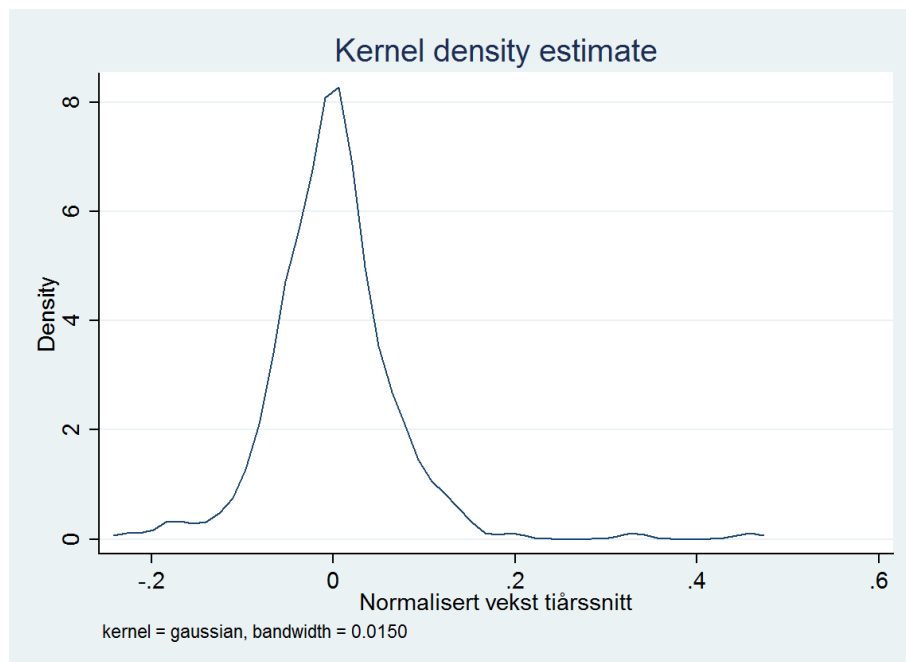
Figur 3: Kernel regresjon med fast vekstsnitt

Som vi ser fra grafen, ligger hovedkonsentrasjonen for den normaliserte vekstraten rundt

-0.05, men ansamles også i en vesentlig grad rundt nullpunktet. Dette hinter om at veksten i relativ stor grad holder seg lik snittet over tid. Men på den andre siden ser vi at hovedtoppen av tetthetsfunksjonen ligger litt under nullpunktet. Man kan derfor påpeke at hovedvekten av vekstraten ligger litt under snittet, men det kan også argumenteres for at dette er marginalt.

Det primære problemet med denne framgangsmåten er at man ikke nødvendigvis tar til rette for teknologisk utvikling, som har vært sterkere periodevis opp igjennom 1800 og 1900 tallet. Dette kan forklare skjevheten vi opplever, med at den normaliserte vekstraten er fokusert rundt -0.05. Ettersom veksten har empirisk blitt sterkere over tid, og denne prosessen kan argumenteres for at er noenlunde treg, vil dette trekke snittet ned. For å korrigere for dette følger vi en endret normalisert vekstrate, hvor snittet vil vokse i forhold til hvilket tiår man ser på. Vi får dermed observert hvordan den normaliserte vekstraten vil holde seg i forhold til snittet, sett ut ifra et kontemporært perspektiv. Dette kan vi argumentere for at inkluderer eventuelle teknologiske fortrinn periodene kan ha ovenfor hverandre

Snittet vil ta formen $\frac{1}{N}\sum g_i$ gitt periode t . Dermed vil en kernel regresjon gi oss følgende estimat:



Figur 4: Kernel regresjon med vekstsnitt som varierer hvert tiår

Som vi ser fra denne grafen ligger hovedvekten av fordelingen til de normaliserte vekstene igjen tilnærmet lik null. Det vil si at topppunktet ligger på tilnærmet lik 0.008. Dette er i større grad lik null enn for den foregående estimasjonen 6, som hadde sin hovedvekt rundt -0.05. Dette insinuerer at veksten i stor grad er lik gjennomsnittsveksten, som er det teorien rundt Gibrats lov tilsier. I tillegg ser vi at fordelingen rundt null er relativt

smal, som også vil forsterke konklusjonen om at norske urbane områder følger Gibrats lov.

Etter å ha sett både den parametriske og den ikke-parametriske analysen av Gibrats lov, kan vi konkludere med at vi ikke kan forkaste hypotesen om Gibrats lov, gitt at vi har et stort nok konfidensintervall. Dette forteller oss at veksten er i all hovedsak uavhengig av tidligere størrelser.

7 Analyse av Zipfs Lov

Modellen som estimeres er av typen¹⁶ $\ln R_{it} = \alpha - \zeta \ln(S_{it})$, hvor man estimerer etter hvert tiår. Resultatene er som gitt i tabellene 11 og 15, hvor vi ser at veksten tenderer mot -1 over 186 år. Resultatene 11 oppnås ved å ta en between-effects-regresjon av modellen, hvor resultatene fra tabell 15 hovedsakelig kommer fra en maximum-likelihood-regresjon. Fra Soo(Tong Soo, 2002) får vi oppgitt at maximum-likelihood-estimatoren for Zipfs lov generelt sett kan oppgis som Hill estimatoren.

Resultatene fra between-effects estimasjonen gir oss en beta-koeffisient som går mot -1. Men konvergens kan finne sted over lang tid. Dette skjer dersom variansen for byene er tilnærmet null, som vil føre til at byene ikke konvergerer. Alternativt vil de konvergere mot Zipfs lov, men i en lav nok hastighet til at prosessen på grensen vil strekke seg til det uendelige, som vist fra Xavier Gabaix (Gabaix, 1999b). Dersom dette ikke er tilfellet vil lengden på prosessen avhengige av variansens størrelse.

I tabell 11 ser vi at paretopparameteren starter på tilnærmet -0.73, og avsluttes på tilnærmet -0.89. Verdt å merke seg er perioden mellom 1900 og 1950, hvor parameteren holder seg stabilt i -0.84 til -0.86 intervallet. Som vi ser fra tabellen med variansen til paretopparameteren, reduseres denne gjennom hele tidsperspektivet vårt. Unntaket ligger i perioden mellom 1910 og 1920, hvor vi får en liten økning. Dette tyder på at hastigheten til trendene reduseres, som tyder på at agglomerasjonsprosessens effekt reduseres noe.

Som vi ser fra tabell 11, er det konvergens-trender fram til 1890. De neste 50 årene er det generelt en divergent trend, med unntak av i perioden 1900 til 1910, hvor et generelt konvergent mønster opptrer fra 1950 til 2001.

Hill estimatorene ser vi som nevnt fra tabell 15, hvor estimatene for $\hat{\zeta}$ er identiske med between-effects estimasjonene. Men som vi ser her er variansene noe lavere enn ved between-effects metoden. Dette indikerer at minste-kvadraters-metode(MKM)-estimasjonen er korrekt, og generelt en noe foretrukket metode. Grunnen for foretrukketheten, er at maximum-likelihood foretrekkes når modellen er ikke-lineær (M. Woolridge, 2009). Et annet argument for dette er at generelt sett vil Hill estimatoren være mer korrekt enn MKM metoden, og når estimatorene til ζ er identiske tilsier dette at standard MKM er å foretrekke.

Resultatene fra begge disse estimatene er skjeve, som vist tidligere. Gitt at vi tar til rette for skjevhetene, det vil si ved å konstruere Monte Carlo beviste korreksjoner instruert fra Ibragimov og Gabaix (Gabaix and Ibragimov, 2006) vil resultatene gis ved tabellene 13 og 17.

De korrigerte between-effects-estimatene opptrer innenfor intervallet [-0.87 , -1.076], når man har tatt hensyn til skjevhetene. Igjen eksisterer det perioder med både konvergens

og divergens, ut ifra endringene til paretoeksponenten. De lengste periodene hvor dette opptrer er i perioden 1835 til 1890, hvor det er konvergens, og perioden 1910 til 1950 som har en divergerende befolkning. Periodene utenom dette kategoriseres som variable i trendene. En annen ting man burde notere som sterk endring mellom standard between-effects-estimering og den robuste estimasjonen er hvordan standardavviket endres mellom de to. Spennet hvor den robuste estimasjonen endres er blitt noe større, som også tilsier at reduksjonen i standardavviket fra 1815 til 2001 har blitt større. Endringene ligger på ca. 0.0825357.

Dersom vi ser på Hill estimatene ser vi at også her går det i perioder med konvergens og divergens. Den mest prominente perioden er fra 1865 til 1890, hvor byene opplevde en sterk konvergens. Deretter følger perioden 1950 til 1980, som er en konvergent periode. Intervallet for $\hat{\zeta}$ ligger mellom -0.87 og -1.076 Standardavviket reduseres her i litt mindre grad enn det gjorde ved de robuste between-effekts-estimatene, men i større grad enn det som var tilfellet i de vanlige maximum likelihood estimatene.

Soo (2002) fant fra sine estimasjoner at paretoeksponenten har forskjellige intervaller, ut fra hvordan man definerer en by (Gabaix and M. Ioannides, 2003). Dersom man kun tar de spesifikke byområdene, for eksempel Oslo med kun 512 589 innbyggere i 2001 (Fosse et al., 2002), så vil eksponentene ha et annet intervall enn om man tar de urbane områdene slik vi har gjort (Tong Soo, 2002). Her vil eksempelet med Oslo ha 1 259 929 innbyggere, og ifølge Rosen og Resnik (1980) falt samtlige land i undersøkelsen deres mellom 0.8 og 1.96 (Tong Soo, 2002). Snittet i denne undersøkelsen var på 1.13, med et standardavvik på 0.19 (Gabaix and M. Ioannides, 2003). Undersøkelsen til Soo fra 2002, som hadde et datasett på 75 forskjellige nasjoners befolkningsundersøkelser, gir oss et snitt for paretoeksponenten på 1.11, som er nokså likt analysen til Rosen og Resnik.

For de byspesifikke undersøkelsene vil estimatene ligge noe høyere enn for de område spesifiserte. Soo (2003) mener her at de byspesifikke parameterene vil ligge på gjennomsnittlig 1.105 mens de område spesifikke parameterene vil ligge på 0.85.

Som vi ser fra våre analyser ligger vi litt over snittene når det gjelder de vanlige regresjonene. Disse begynner på tilnærmet 0.7295, helt i begynnelsen av målingene, og avsluttes over snittet med tilnærmet 0.89. Dersom vi ser på estimatene for hele perioden, det vil si estimerer for alle periodene samlet, så ser vi fra tabell 6 at MKM estimatet ligger på 0.829. Dette ligger under snittet både Soo (Tong Soo, 2002) og Rosen og Resnik (Gabaix and M. Ioannides, 2003) kommer fram til. Dersom vi ser på det robuste estimatet, gitt ved tabell 19 ser vi at estimatet ligger litt over snittene vi har sett på. For Norge er eksponenten ved den robuste estimasjonen på 0.995, som er ganske nære den optimale eksponenten på 1.

Tidligere påpekte vi at standardavvikene var noe ukorrekte. Ved å ta ibruk Gabaix og Ibragimov sin korreksjon av disse (Giesen and Suedekum, 2009) vil vi få mer korrekte

standardavvik, som vil gi et mer komplett bilde av hvordan Zipfs lov forholder seg til norske data. Regelen som brukes er gitt ved (Giesen and Suedekum, 2009):

$$\sqrt{\frac{2}{n}\zeta} \quad (44)$$

Igjen kan vi se resultatene i tabellene, 11 - 20, som igjen tillater et noe større spenn på Zipf estimatene. Standardavvikene for både Between Effects og Hill estimatorene er igjen identiske takket være regelen vi bruker⁴⁴. Vi ser at standardavviket er vesentlig høyere enn det man opprinnelig estimerer, som viser til volatiliteten som estimasjonen ikke nødvendigvis får med seg når det er et fåtall av urbane strøk. Med andre ord kommer skjevhetene som oppstår ved mindre datasett fram ved bruk av denne metoden. Standardavviket varierer mellom 0.27 og 0.41 i between-effects-estimatene, der de som er Monte Carlo-justert generelt ligge høyere enn de som ikke er justert.

Ser vi på R^2 , som forklarer tilpasningen til modellen, så ser vi at denne generelt ligger høyt. Det er vanlig at R^2 ligger ganske høyt, ettersom vi ordner dataene våre etter rang. Dette insinuerer at modellen har en sterk forklaringsevne.

Dersom vi velger å teste hypotesen om at $\zeta = 1$ vil dette gi oss et bedre bilde av hvorvidt modellen som er estimert fra våre data passer inn med Zipfs lov. Alternativ hypotese er at ζ er betydelig forskjellig fra 1. Dette kan representeres på følgende måte:

$$H_0 : \zeta = 1 \quad (45)$$

Som er null hypotesen, og:

$$H_1 : \zeta \neq 1 \quad (46)$$

Alternativhypotesen er at det ikke finnes en lineær sammenheng med stigningstall statistisk lik 1. Dette betyr at dersom alternativhypotesen stemmer, vil det være en mye jevnere fordeling mellom byene, alternativt mye skjevere, enn det Zipfs lov tilsier.

En t-test for hvert estimat i tabell 11 gir t-statistikk som hovedsakelig er over den kritiske verdien på 1.96 for 95% konfidensintervallet, samt over 2.576 ved 99% konfidensintervallet. Den minste t-verdien er på 1.95, mens den største ligger på 3.15. Det er én eneste periode som ikke forkaster nullhypotesen fra t-testen ved 95% konfidensintervall, perioden rundt 1890. Ved 99% konfidensintervall er det perioden 1865 til 1930 som ikke forkaster nullhypotesen. Verdiene for t-statistikken er listet opp i tabell 12. Disse t-verdiene gjelder hovedsakelig for den opprinnelige standard feilen, som vi vet at viser en sterk skjevhet. Dermed kan vi anta at selv ved den største skjevheten som modellen kan inneha vil deler

av estimatene ikke kunne forkaste nullhypotesen.

Dersom vi tar i bruk standardfeil korrigert for skjevheten, med formelen 44 vil vi få mer korrekte t-verdier. Her vil verdiene variere mellom 0.3 og 0.99, som man kan se fra tabell 14. Ingen av disse t-verdiene går over kritisk verdi ved 95% konfidensintervall, som betyr at vi ikke kan forkaste nullhypotesen til fordel for alternativ hypotesen. Zipfs lov vil da holde for de norske urbane miljøene, ettersom paretoparameterene ikke er betydelig forskjellig fra 1.

Dersom korreksjonen av standard feilen ikke skulle gi stabile resultater, så vil vi se på Monte Carlo bevist korrigerings av den avhengige variabelen. Dette er gitt ved $\ln(\text{Rank} - 1/2)$, som vist tidligere. T-statistikken vår vil ligge mellom 0.14 og 1.6 for samtlige tester. Dette er ikke over den kritiske verdien ved et 95% konfidensintervall, som fører til at vi ikke kan forkaste null hypotesen. Igjen kan vi ikke motbevise at Zipfs lov gjelder for norske urbane miljøer.

Ved å se på korreksjonen av standardfeilen med endret formuleringen av modellen, så ser vi at t-statistikken vil hovedsakelig ligge i intervallet mellom 0.02 og 0.41. Dette er mindre enn den kritiske verdien, slik at null hypotesen vil ikke kunne forkastes. Zipfs lov vil derfor gjelde for Norges urbane områder.

Ettersom null hypotesen kun kan forkastes ved de skjevste estimatene, og selv der kun kan forkastes utenom perioden 1875 til 1920 ved et 99% konfidensintervall, kan vi anta at Zipfs lov gjelder for Norge. Vi kan da si at paretoparameteren ikke er betydelig forskjellig fra den optimale Zipf-verdien på 1, noe som tilsier en svært lineær sammenheng mellom rang og størrelse på urbane områder.

Dersom vi ser på Hill-estimatorene, det vil si maximum likelihood-estimasjon, ser vi at estimatene vil være identiske med between-effects-estimasjonen. De største avvikene vi ser er mellom standard feilene til de forskjellige estimatene, som vil produsere noe annerledes t-statistikker. Fra 16 ser vi t-statistikken for de ordinære standard avvikene, innenfor intervallet 1.91 og 3.4. Perioden som ikke er i stand til å forkaste null hypotesen ved et 99% konfidensintervall er tidsrommet mellom 1875 og 1920. Som vi ser er dette et noe mindre tidsrom enn ved between effects estimasjonen. Samtidig vil samme periode som tidligere ikke forkaste null hypotesen ved det noe mindre 95% konfidens intervallet. 1890 har den laveste t-verdien av dette utvalget, med en verdi på 1.91.

Ved bruken av Gabaix-Ibragimov korreksjonen av standardavviket, vil samtlige estimasjoner av ζ være så lik 1 at forskjellene ikke er statistisk signifikante. Med andre ord vil alle t-verdiene ligge under den kritiske verdien på 1.96. T-verdiene er listet hos tabell 16.

Dersom vi bruker Monte Carlo-bevist korreksjon av modell-spesifikasjonen, vil t-statistikken ligge innenfor intervallet 0.16 og 1.7118, som er under den kritiske verdien på 1.96. Ved å

korrigere standardavvikene i henhold til Gabaix-Ibragimov, vil dette intervallet reduseres ytterligere til verdier mellom 0.027 og 0.40818.

Vi kan ikke forkaste null hypotesen for den robuste Hill-estimasjonen, selv med skjeve ordinære standardavvik. Som ved between-effects-estimatene kan vi kun forkaste nullhypotesen for estimatene med størst skjevhet, og selv der er det perioder hvor man ikke kan forkaste nullhypotesen.

Avslutningsvis vil vi ta totale estimater av ligning 38 over alle tidsobservasjonene. Ettersom vi vet at den mest robuste estimasjonsformen kommer fra Gabaix-Ibragimov korrigert modell spesifikasjon, med Gabaix-Ibragimov korreksjon av standardavvikene (Gabaix and Ibragimov, 2006).

Resultatet av en between effects estimasjon av 38 kan observeres ved tabell 19. Estimatet ligger på 0.995, som er relativt nære 1. En Hill estimasjon av ligning 38 gir et estimat på 0.568. Dette er relativt langt unna den optimale pareto parameteren. Dersom vi tar t-tester på disse to estimatene ser vi at vi generelt ikke kan forkaste null hypotesen for between-effects-estimatene, mens vi kan forkaste null hypotesen for Hill-estimatoren. T-verdiene er henholdsvis 0.0578 og 8.77 for between-effects og Hill-estimatorene.

Dette insinuerer at Hill metoden ikke er optimal for å estimere modellen over alle tidsobservasjonene. Når vi har estimert over de individuelle tidsobservasjonene, er den eneste forskjellen mellom between-effects og Hill-estimasjonene vært standardfeilene. Men her er det store forskjeller mellom de to estimasjonsformene. Dette er mest sannsynlig på grunn av den lineære naturen til dataene, ettersom de er ordnet på en spesifikk måte. Maximum likelihood vil generelt være best spesifisert ovenfor dynamiske data.

8 Konklusjon

Fra Gabaix og Ioannides (2003) påpekes det at hovedspørsmålet ved empirisk arbeid ikke burde være hvorvidt dataene passer en teori perfekt, men i hvor stor grad teorien passer dataene. (Gabaix and M. Ioannides, 2003). Hovedhensikten med denne oppgaven var å undersøke om hvorvidt de urbane miljøene i Norge følger regularitetene som Gibrat (1931) og Zipf (1949) presenterer, og om mulig gi en forklaring på hvordan dette går til.

Våre undersøkelser viser til at regularitetene vi undersøker, Zipfs og Gibrats lov, i relativ stor grad vil passe dataene vi undersøker. Generelt kunne vi forkaste Gibrats lov, men dette gjaldt ikke 99% konfidens intervallet, noe som henter til at veksten i større eller mindre grad følger prinsippet bak Zipfs lov. Vi kan anta at det enten er en veldig svakt korrelasjon med tidligere størrelser, eller at det kan være andre faktorer som virker inn på vekst enn de alternativene vi bruker i modellen. Krugman (Krugman, 1996) argumenterer for at negative og positive eksternaliteter som et resultat av agglomerasjon er stedsbundet, gjerne i form av industriformen det urbane området spesialiserte seg på.

Vår analyse tar ikke for seg eventuelle negative eksternaliteter som kan påvirke evolusjonen til et urbant område, så at dette kan være former for uobserverte faktorer som vil påvirke vekst over tid. I tillegg har vi forsøkt å ta hensyn til teknologisk vekst, ved å ta tiårs spesifikt vekstsnitt. Dette er en av de enkleste metodene vi kan ta i bruk, men den vil i mindre grad ta for seg hvordan teknologiske framspring påvirker de enkelte byene. På den andre siden får vi et visst bilde av hvordan den generelle utviklingen har vært.

Når vi ser på Zipfs lov gir dette et litt annet bilde. Her avhenger konklusjonen vår av hvilken estimasjonsmetode og modellspesifikasjon som er brukt. Med de som har størst forventningsskjevhet vil vi kun i enkelte perioder ikke kunne forkaste nullhypotesen. Samtidig vil spesifikasjonene som gir minst forventningsskjevhet ikke kunne forkaste nullhypotesen i det hele tatt. Dette gir oss et inntrykk om at det eksisterer en lineær sammenheng mellom rang og størrelse. Spørsmålet her ligger i hvorfor det eksisterer en så spesifikk sammenheng mellom rangen et urbant område innehar, og størrelsen til nevnte urbane område. Fra Gabaix (Gabaix, 1999b) utledes regulariteten som en stokastisk prosess. Dette er forklart i teorikapitlet, men denne framgangsmåten gir ikke noen økonomiske mekanismer til grunn for selve fordelingen. Som vi nevnte i henhold til konklusjonen for Gibrats lov kan vekst være avhengig av eksternaliteter. Dette kan bety at, for Zipfs lov, mengden eksternaliteter vil være med på å bestemme hvor et urbant område vil ende opp i distribusjonen.

Vi kan også se fra paretoparameteren som estimeres i Zipfs lov hvordan bosetningsmønsteret er. Den generelle trenden vi kan observere fra tabellene 11, 13, 15, og 17 er konvergens. Med andre ord vil det i større grad være flere som flytter inn i de urbane områdene, og vi får negativ befolkningsvekst i landslige strøk. Edward Glaeser påpeker potensialet som

8 KONKLUSJON

byer representererL. Glaeser (2011), som kan være en årsakene til avbefolkning av rurale områder. Men ettersom det per i dag ikke eksisterer en god nok modell som både gir et godt bilde av agglomerasjonsmekanismene, og samtidig ikke ofrer for mye av realismen til antagelser, blir det med spekulasjoner.

Referanser

- (2002-2003). Stortingsmeldingen - om utviklingen av storbypolitikk. Research Paper.
- (2008). United nations expert group meeting on population distribution, urbanization, internal migrations and development. Research Paper.
- Black, D. and V. Henderson (2003). Urban evolution in the usa. *Journal of Economic Geography* 3(4), 343–372.
- Clarke, J. S. and J. C. Stabler (1991). Gibrat’s law and the growth of canadian cities. *Urban Studies* 28(4), 635–639.
- Eeckhout, J. (2004). Gibrat’s law for (all) cities. *American Economic Review* (94), 1429–1451.
- Fosse, J. K., J. Karlsen, M.-L. Magnussen, and J. Cruickshank (2002). Norske byregioner - en drøftning av byenes geografiske utbredelse. Research Paper.
- G. Overman, H. and Y. Ioannides (2000). Zipfs law for cities: An empirical examination. Centre for Economic Performance.
- Gabaix, X. (1999a). Zipf’s law and the growth of cities. *The American Economic Review* 89(2), 129–132.
- Gabaix, X. (1999b). Zipf’s law for cities: An explanation. *The Quarterly Journal of Economics* 114(3), 739–767.
- Gabaix, X. and R. Ibragimov (2006). Log(rank-1/2): A simple way to improve the ols estimation of tail exponents. Research Paper.
- Gabaix, X. and Y. M. Ioannides (2003). The evolution of city size distributions. *Handbook of Urban and Regional Economics Vol. IV: Cities and Geography* (IV), 2–51.
- Giesen, K. and J. Suedekum (2009). Zipf’s law for cities in the regions and the country. Research Paper.
- I. Severide, P. and S. Longa (2004). Folke- og boligtelling 2001. Statistisk Sentralbyrå.
- J. Bogue, D. (1953). Population growth in standard metropolitan areas 1900 - 1950. Research Paper.
- Juvkam, D. (1999). Historisk oversikt over endringer i kommune- og fylkesinndeling. Research Paper.
- Krugman, P. (1996). Confronting the mystery of urban hierarchy. *Journal of the Japanese and International Economics* 10, 399–418.

REFERANSER

- L. Glaeser, E. (2011). Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier.
- Lanaspa, R. G.-V. L. and F. Sanz (2010). Gibrat's law for cities revisited. University of Zaragoza.
- Le Gallo, J. and C. Chasco (2008). Spatial analysis of urban growth in Spain, 1900-2001. *Empirical Econometrics* (34), 59–80.
- M. Wooldridge, J. (2009). *Introductory Econometrics - A Modern Approach*.
- Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* (22), 265–312.
- Reiss, R.-D. (1989). *Approximate distribution of order statistics*. Springer Verlag.
- Soltvedt, K. (2004). *Folketellinger gjennom 200 år*. Kongsvinger: Statistisk sentralbyrå.
- Tong Soo, K. (2002). Zipf's law for cities: A cross country investigation. Research Paper.
- Verbeek, M. (2008). *A guide to modern econometrics*.

A Tabeller

A.1 Tabeller for deskriptiv statistikk

Tabell 1: 1815

Rank	Region	Folkemengde	F/F_{Oslo}	F/F_{snitt}	F^{Rank}/F^{Rank-1}
1	Oslo	114232	1	2,996	-
2	Bergen	53120	2,15	1,393	0,465
3	Mjøsbyene	47806	2,389	1,254	0,9
4	Trondheim	47792	2,39	1,253	1
5	Kristiansand	28954	3,945	0,759	0,606
6	Grenland	20231	5,646	0,531	0,699
7	Stavanger	18918	6,038	0,496	0,935
8	Tønsberg	18206	6,274	0,477	0,962
9	Glomma	16932	6,747	0,444	0,93
10	Larvik	15108	7,561	0,396	0,892

Tabell 2: 1875

Rank	Region	Folkemengde	F/F_{Oslo}	F/F_{snitt}	F^{Rank}/F^{Rank-1}
1	Oslo	289882	1	3,6	-
2	Bergen	91895	3,154	1,141	0,317
3	Mjøsbyene	80316	3,609	0,997	0,874
4	Trondheim	75531	3,838	0,938	1
5	Stavanger	55857	5,19	0,694	0,606
6	Kristiansand	52616	5,509	0,653	0,699
7	Glomma	50048	5,792	0,622	0,935
8	Tønsberg	41901	6,918	0,52	0,962
9	Grenland	35665	8,128	0,443	0,93
10	Larvik	31466	9,213	0,391	0,892

A TABELLER

Tabell 3: 1910

Rank	Region	Folkemengde	F/F_{Oslo}	F/F_{snitt}	F^{Rank}/F^{Rank-1}
1	Oslo	493167	1	4,081	-
2	Bergen	151998	3,245	1,273	0,308
3	Trondheim	99697	4,947	0,830	0,656
4	Mjøsbyene	96008	5,137	0,799	0,963
5	Glomma	81826	6,027	0,684	0,852
6	Stavanger	78754	6,262	0,656	0,962
7	Kristiansand	57216	8,619	0,476	0,727
8	Grenland	54350	9,074	0,452	0,95
9	Tønsberg	51359	9,602	0,428	0,945
10	Larvik	38625	12,768	0,322	0,752

Tabell 4: 1950

Rank	Region	Folkemengde	F/F_{Oslo}	F/F_{snitt}	F^{Rank}/F^{Rank-1}
1	Oslo	788391	1	4,367	-
2	Bergen	224823	3,507	1,262	0,285
3	Trondheim	143424	5,497	0,805	0,638
4	Mjøsbyene	136879	5,76	0,768	0,954
5	Stavanger	127632	6,177	0,717	0,932
6	Glomma	106671	7,391	0,548	0,863
7	Grenland	73315	10,753	0,412	0,687
8	Kristiansand	73119	10,782	0,410	0,997
9	Tønsberg	70401	11,199	0,395	0,963
10	Larvik	56286	14,007	0,316	0,8

Tabell 5: 2001

Rank	Region	Folkemengde	F/F_{Oslo}	F/F_{snitt}	F^{Rank}/F^{Rank-1}
1	Oslo	1264003	1	4,4	-
2	Bergen	338137	3,738	1,216	0,268
3	Stavanger	261074	4,842	0,939	0,772
4	Trondheim	225455	5,606	0,81	0,864
5	Mjøsbyene	174267	7,253	0,626	0,773
6	Kristiansand	135162	9,352	0,486	0,776
7	Glomma	127683	9,9	0,459	0,945
8	Tønsberg	106693	11,847	0,384	0,836
9	Grenland	106222	11,9	0,382	0,996
10	Larvik	83226	15,188	0,3	0,784

A.2 Tabeller for økonometriske spesifikasjoner

Tabell 6: Between Effects estimasjon

Variabel	Koeffisient	σ_ζ
Infolkemengde	-0.829	(0.059)
Intercept	10.868	(0.643)

B Regresjons Resultater

B.1 Gibrat estimasjoner

Tabell 7: Random Effects

Variabel	Koeffisient	(σ)
lnforrigefolke	-0.0000000526	(0.0000000276)
Intercept	0.1180296	(0.0050663)

Tabell 8: Robust Random Effects

Variabel	Koeffisient	(σ)
lnfolkeforrige	0.0000000526	(0.0000000212)
Intercept	0.1180296	(0.0051062)

Tabell 9: Fixed Effects

Variabel	Koeffisient	(σ)
lnforrigefolke	-0.0000000496	(0.0000000286)
Intercept	0.1177393	(0.0051856)

Tabell 10: Robust Fixed Effects

Variabel	Koeffisient	σ
lnfolkeforrige	0.0000000496	(0.0000000238)
Intercept	0.1177393	(0.0023251)

B.2 Zipf estimasjoner

Tabell 11: BE estimasjon

Year	$\hat{\alpha}$	$\sigma_{\hat{\alpha}}$	$\hat{\zeta}$	$\sigma_{\hat{\zeta}}$	R^2	GI-korrigert $\sigma_{\hat{\zeta}}$
1815	8.969916	(0.9868071)	0.7294781	(0.099068)	0.8313	(0.2757)
1825	9.0868071	(0.9868071)	0.7319198	(0.0853662)	0.8597	(0.2766)
1835	9.192523	(0.8821327)	0.7311764	(0.0868642)	0.8552	(0.2764)
1845	9.335824	(0.9280888)	0.7366067	(0.0903359)	0.8471	(0.2784)
1855	9.642501	(0.9268168)	0.7558685	(0.0889754)	0.8574	(0.2857)
1865	9.885658	(0.9794223)	0.767156	(0.0925836)	0.8512	(0.2899)
1875	10.4175	(0.9376661)	0.8096017	(0.0877977)	0.8763	(0.306)
1890	11.09027	(0.8452617)	0.8613019	(0.0781296)	0.9101	(0.3255)
1900	11.0662	(0.7252671)	0.8482852	(0.066191)	0.9319	(0.3206)
1910	11.19156	(0.7104933)	0.8530675	(0.0643421)	0.9361	(0.3224)
1920	11.20187	(0.7283966)	0.8443626	(0.0652198)	0.9332	(0.319)
1930	11.21825	(0.7033624)	0.8412469	(0.0626365)	0.9376	(0.318)
1946	11.29112	(0.6528857)	0.8386598	(0.0575186)	0.9466	(0.317)
1950	11.36036	(0.6506795)	0.8413796	(0.0570958)	0.9476	(0.318)
1960	11.65761	(0.6492058)	0.8582383	(0.0563384)	0.9508	(0.3244)
1970	11.84929	(0.6893218)	0.8652611	(0.0591935)	0.9468	(0.327)
1980	12.05783	(0.6467888)	0.8765311	(0.0551254)	0.9547	(0.3313)
1990	12.26429	(0.5608017)	0.8906931	(0.047614)	0.9668	(0.3367)
2001	12.3386	(0.5202628)	0.8903062	(0.0438426)	0.9717	(0.3365)

B REGRESJONS RESULTATER

Tabell 12: t-statistikk for between effects estimasjon

År	ζ	t-statistikk	GI. Korr. $\sigma_{\hat{\zeta}}$	t-statistikk GI. Korr. $\sigma_{\hat{\zeta}}$	
1815	0.7294781	0.0990680	-2.7306688	0.2757	0.9812184
1825	0.7319198	0.0853662	-3.1403553	0.2766	0.9691981
1835	0.7311764	0.0868642	-3.0947571	0.2764	0.9725890
1845	0.7366067	0.0903359	-2.9157101	0.2784	0.9460966
1855	0.7558685	0.0889754	-2.7438090	0.2857	0.8545030
1865	0.7671560	0.0914736	-2.5454776	0.2899	0.8031873
1875	0.8096017	0.0877977	-2.1686024	0.3060	0.6222167
1890	0.8613019	0.0712960	-1.9453840	0.3255	0.4261078
1900	0.8482852	0.0661910	-2.2920760	0.3206	0.4732215
1910	0.8530675	0.0643421	-2.2836137	0.3224	0.4557460
1920	0.8443626	0.0652198	-2.3863520	0.3190	0.4878915
1930	0.8412469	0.0626365	-2.5345142	0.3180	0.4992236
1946	0.8386598	0.0575186	-2.8050092	0.3170	0.5089596
1950	0.8413769	0.0570958	-2.7781921	0.3180	0.4988148
1960	0.8582383	0.0563384	-2.5162536	0.3244	0.4369966
1970	0.8652611	0.0591935	-2.2762449	0.3270	0.4120456
1980	0.8765311	0.0551254	-2.2397824	0.3313	0.3726800
1990	0.8906931	0.0476140	-2.2956882	0.3367	0.3246418
2001	0.8903062	0.0438426	-2.5019912	0.3365	0.3259845

Tabell 13: Robust(Monte Carlo) BE estimasjon

Year	$\hat{\alpha}$	$\sigma_{\hat{\alpha}}$	$\hat{\zeta}$	$\sigma_{\hat{\zeta}}$	R^2	GI-korrigert $\sigma_{\hat{\zeta}}$
1815	10.21991	(1.297977)	0.8700077	(0.130307)	0.8021	(0.3288)
1825	10.31118	(1.120449)	0.8683676	(0.1120004)	0.8336	(0.3282)
1835	10.42602	(1.1594)	0.8666588	(0.114167)	0.8276	(0.3276)
1845	10.60353	(1.209037)	0.8738444	(0.1158829)	0.8213	(0.3303)
1855	10.9793	(1.2071)	0.8978466	(0.1158829)	0.8334	(0.3394)
1865	11.28801	(1.258443)	0.9131408	(0.1189591)	0.8308	(0.3451)
1875	11.94904	(1.202118)	0.9662921	(0.1125594)	0.8600	(0.3652)
1890	12.81456	(1.202118)	1.033796	(0.0976721)	0.9032	(0.3908)
1900	12.80806	(0.8948816)	1.020222	(0.0816707)	0.9286	(0.3856)
1910	12.95318	(0.8841525)	1.025461	(0.0800687)	0.9318	(0.3876)
1920	12.97225	(0.8980919)	1.015597	(0.0804141)	0.9300	(0.3839)
1930	12.99744	(0.8626946)	1.012339	(0.0768255)	0.9354	(0.3826)
1946	13.09804	(0.787746)	1.010366	(0.0693997)	0.9464	(0.3819)
1950	13.18155	(0.784994)	1.013651	(0.0688817)	0.9475	(0.3831)
1960	13.5641	(0.758981)	1.035598	(0.0658648)	0.9537	(0.3914)
1970	13.8054	(0.7879342)	1.045419	(0.0676616)	0.9521	(0.3951)
1980	14.03133	(0.7677577)	1.056812	(0.0654355)	0.9560	(0.3994)
1990	14.29779	(0.6331725)	1.075379	(0.0537585)	0.9709	(0.4064)
2001	14.39461	(0.5668831)	1.075512	(0.0477713)	0.9769	(0.4065)

B REGRESJONS RESULTATER

Tabell 14: t-statistikk for robust between effects estimasjon

År	ζ	t-statistikk	GI. Korr. $\sigma_{\hat{\zeta}}$	t-statistikk GI. Korr. $\sigma_{\hat{\zeta}}$	
1815	0.8700077	0.1303070	0.9975849	0.3288	0.3953537
1825	0.8683676	0.1120004	1.1752851	0.3282	0.4010737
1835	0.8666588	0.1141670	1.1679487	0.3276	0.4070244
1845	0.8738444	0.1158290	1.0891538	0.3303	0.3819425
1855	0.8978466	0.1158829	0.8815226	0.3394	0.3009823
1865	0.9131408	0.1189591	0.7301602	0.3451	0.2516928
1875	0.9662921	0.1125594	0.2994677	0.3652	0.0922998
1890	1.0337960	0.0976721	0.3460149	0.3908	0.0864790
1900	1.0202220	0.0816707	0.2476041	0.3856	0.0524429
1910	1.0254610	0.0800687	0.3179894	0.3876	0.0656889
1920	1.0155970	0.0804141	0.1939585	0.3839	0.0406278
1930	1.0123390	0.0768255	0.1606107	0.3826	0.0322504
1946	1.0103660	0.0693997	0.1493666	0.3819	0.0271432
1950	1.0136510	0.0688817	0.1981804	0.3831	0.0356330
1960	1.0355980	0.0658648	0.5404708	0.3914	0.0909504
1970	1.0454190	0.0676616	0.6712670	0.3951	0.1149557
1980	1.0568120	0.0654355	0.8682137	0.3999	0.1420513
1990	1.0753790	0.0537585	1.4021783	0.4064	0.1854798
2001	1.0755120	0.0477713	1.5806980	0.4065	0.1857614

Tabell 15: Hill estimasjon

Year	$\hat{\alpha}$	$\sigma_{\hat{\alpha}}$	$\hat{\zeta}$	$\sigma_{\hat{\zeta}}$	GI-korrigert $\sigma_{\hat{\zeta}}$
1815	8.969916	(0.9077276)	0.7294781	(0.091129)	(0.2757)
1825	9.088684	(0.7906504)	0.7319198	(0.0790336)	(0.2766)
1835	9.192523	(0.8166949)	0.7311764	(0.0804205)	(0.2764)
1845	9.335824	(0.8592419)	0.7366067	(0.0836346)	(0.2784)
1855	9.642501	(0.8580643)	0.7558685	(0.0823751)	(0.2857)
1865	9.885658	(0.9067674)	0.767156	(0.0857156)	(0.2899)
1875	10.4175	(0.8681087)	0.8096017	(0.0812847)	(0.306)
1890	11.09027	(0.782559)	0.8613019	(0.0723339)	(0.3255)
1900	11.0662	(0.6714658)	0.8482852	(0.0612809)	(0.3206)
1910	11.19156	(0.6577879)	0.8530675	(0.0595691)	(0.3224)
1920	11.20187	(0.6743631)	0.8443626	(0.06063817)	(0.319)
1930	11.21825	(0.651186)	0.8412469	(0.05799)	(0.318)
1946	11.29112	(0.6044537)	0.8386598	(0.0532518)	(0.317)
1950	11.36036	(0.6024112)	0.8413796	(0.0528604)	(0.318)
1960	11.66228	(0.6010468)	0.8582383	(0.0521592)	(0.3244)
1970	11.84929	(0.638187)	0.8652611	(0.0548025)	(0.327)
1980	12.05783	(0.5988091)	0.8765311	(0.0510361)	(0.3313)
1990	12.26429	(0.5192007)	0.8906931	(0.0440819)	(0.3367)
2001	12.3386	(0.481669)	0.8903062	(0.0405903)	(0.3365)

B REGRESJONS RESULTATER

Tabell 16: t-statistikk for Hill estimatorene

År	ζ	t-statistikk	GI. Korr. $\sigma_{\hat{\zeta}}$	t-statistikk GI. Korr. $\sigma_{\hat{\zeta}}$	
1815	0.7294781	0.0911290	2.9685600	0.2757	0.9812184
1825	0.7319198	0.0790336	3.3919776	0.2766	0.9691981
1835	0.7311764	0.0804205	3.3427248	0.2764	0.9725890
1845	0.7366067	0.0836346	3.1493341	0.2784	0.9460966
1855	0.7558685	0.0823751	2.9636565	0.2857	0.8545030
1865	0.7671560	0.0857156	2.7164717	0.2899	0.8031873
1875	0.8096017	0.0812847	2.3423633	0.3060	0.6222167
1890	0.8613019	0.0723339	1.9174702	0.3255	0.4261078
1900	0.8482852	0.0612809	2.4757273	0.3206	0.4732215
1910	0.8530675	0.0595691	2.4665892	0.3224	0.4557460
1920	0.8443626	0.0606382	2.5666573	0.3190	0.4878915
1930	0.8412469	0.0579900	2.7375944	0.3180	0.4992236
1946	0.8386598	0.0532518	3.0297605	0.3170	0.5089596
1950	0.8413769	0.0528604	3.0007927	0.3180	0.4988148
1960	0.8582383	0.0521592	2.7178657	0.3244	0.4369966
1970	0.8652611	0.0548025	2.4586269	0.3270	0.4120456
1980	0.8765311	0.0510361	2.4192464	0.3313	0.3726800
1990	0.8906931	0.0440819	2.4796322	0.3367	0.3246418
2001	0.8903062	0.0405903	2.7024634	0.3365	0.3259845

Tabell 17: Robust Hill estimasjon

Year	$\hat{\alpha}$	$\sigma_{\hat{\alpha}}$	$\hat{\zeta}$	$\sigma_{\hat{\zeta}}$	GI-korrigert $\sigma_{\hat{\zeta}}$
1815	10.21991	(1.193961)	0.8700077	(0.1198647)	(0.3288)
1825	10.31118	(1.037332)	0.8683676	(0.103692)	(0.3282)
1835	10.42602	(1.073394)	0.8666588	(0.1056979)	(0.3276)
1845	10.60353	(1.119349)	0.8738444	(0.1089523)	(0.3303)
1855	10.9793	(1.117556)	0.8978466	(0.1072865)	(0.3393)
1865	11.28801	(1.16509)	0.9131408	(0.1101346)	(0.3445)
1875	11.94904	(1.112943)	0.9662921	(0.1042096)	(0.3652)
1890	12.81456	(0.9783)	1.033796	(0.0904267)	(0.391)
1900	12.80806	(0.8284981)	1.020222	(0.0756123)	(0.3856)
1910	12.95318	(0.8185649)	1.025461	(0.0741291)	(0.3876)
1920	12.97225	(0.8314702)	1.015597	(0.0744489)	(0.3839)
1930	12.99744	(0.7986987)	1.012339	(0.0711265)	(0.3826)
1946	13.09804	(0.7293099)	1.010366	(0.0642515)	(0.3819)
1950	13.18155	(0.7267621)	1.013651	(0.0637719)	(0.3828)
1960	13.5641	(0.7026788)	1.035598	(0.0609788)	(0.3914)
1970	13.8054	(0.7294842)	1.045419	(0.0626423)	(0.395)
1980	14.03133	(0.7108044)	1.056812	(0.0605814)	(0.3994)
1990	14.29799	(0.5862029)	1.075379	(0.0497707)	(0.4064)
2001	14.39461	(0.5248309)	1.075512	(0.0442275)	(0.4065)

B REGRESJONS RESULTATER

Tabell 18: t-statistikk for de robust (Monte Carlo) Hill estimatorene

År	ζ	t-statistikk	GI. Korr. $\sigma_{\hat{\zeta}}$	t-statistikk GI. Korr. $\sigma_{\hat{\zeta}}$	
1815	0.8700077	0.1198647	1.0844919	0.3288	0.3953537
1825	0.8683676	0.1036920	1.2694557	0.3282	0.4010737
1835	0.8666588	0.1056979	1.2615312	0.3276	0.4070244
1845	0.8738444	0.1089523	1.1578975	0.3303	0.3819425
1855	0.8978466	0.1072865	0.9521552	0.3394	0.3009825
1865	0.9131408	0.1101346	0.7886641	0.3451	0.2516928
1875	0.9662921	0.1042096	0.3234625	0.3652	0.0922998
1890	1.0337960	0.0904267	0.3737392	0.3908	0.0864790
1900	1.0202220	0.0756123	0.2674433	0.3856	0.0524429
1910	1.0254610	0.0741291	0.3434684	0.3876	0.0656889
1920	1.0155970	0.0744489	0.2094994	0.3839	0.0406278
1930	1.0123390	0.0711265	0.1734796	0.3826	0.0322504
1946	1.0103660	0.0642515	0.1613348	0.3819	0.0271432
1950	1.0136510	0.0637719	0.2140598	0.3831	0.0356330
1960	1.0355980	0.0609788	0.5837767	0.3914	0.0909504
1970	1.0454190	0.0626423	0.7250532	0.3951	0.1149557
1980	1.0568120	0.0605814	0.9377796	0.3999	0.1420513
1990	1.0753790	0.0497707	1.5145256	0.4064	0.1854798
2001	1.0755120	0.0442275	1.7073540	0.4065	0.1857614

B REGRESJONS RESULTATER

Tabell 19: Robust Between Effects estimator

c c	Variable	Coefficient	(Std. Err.)	(GI. korrigert $\sigma_{\hat{\zeta}}$)
	lnfolkemengde	0.995	(0.078)	(0.8644)
	Intercept	12.548	(0.853)	(-)

Med t-statistikker på 0.064 for den ordinære standard feilen, og 0.0578 for den korrigerte standard feilen.

Tabell 20: Robust Hill Estimator, total

Variable	Coefficient	(Std. Err.)
Equation 1 : lnrankMC		
lnfolkemengde	0.568	(0.016)
Intercept	8.831	(0.000)
Equation 2 : sigma_u		
Intercept	0.637	(0.000)
Equation 3 : sigma_e		
Intercept	0.394	(0.020)

T-statistikken for den ordinære standard feilen er gitt som 27. Den GI-korrigert $\sigma_{\hat{\zeta}}$ for Hill estimaten er (0.0493), som fører til en t-statistikk på 8.77.